Text Data Compression Software Design Using Arithmetic Coding Method

Al Aris

STMIK Budidarma, Indonesia

Email : Alaris@gmail.com

Received : 12 April 2020

Revised : 20 Mei 2020

Accepted : 28 Mei 2020

© 2020 The Author: Published by. Cattleya Darmaya Fortuna

Abstract

Data compression is the science or art of representing information in a more compact form. Various algorithms have been developed for data compression purposes. One of the algorithms used is Arithmetic Coding. The working principle of this algorithm is to replace a series of input symbols with floating point numbers, where the longer and more complex the encoded message, the more bits required for file compression. This software is designed using the Visual Basic 6.0 programming language and is able to compress text files with a compression rate of about 75 percent. This system can also be used to compress two or more text files at once.

Keywords: Text Data Compression, Arithmetic Coding Method

INTRODUCTION

Technological developments are getting faster and faster, resulting in a lot of digital data stored in a storage media. Especially if the data is continuously stored in an unspecified time. So the amount of storage media needed to accommodate the data [1]. Data is something that has no meaning for the recipient and still requires processing. Data can be in the form of a situation, images, sounds, letters, numbers, mathematics, language or other symbols that we can use as material to see the environment, objects, events or concepts. The size of the data that must be sent exceeds the transmission speed of the existing hardware, so there is still a relatively large delay time. In addition to storage media such as floppy disks, hard disks, and compact disks have limited capacity. There are many theories and algorithms for data compression, including the Huffman method, Run-Length Encoding (RLE), Lempel-Zip-Welch (LZW), Shannon-Fano, and several other algorithms [2]. Information is a very valuable thing. When information or data falls into irresponsible hands, it will be a disaster for the owner. There are various ways to protect data or information. In exchanging data and information, a high level of security is needed, this is due to the frequent occurrence of data theft [3]. Data or information security techniques have developed very rapidly. The process of securing messages that are widely circulated includes using cryptography and steganography through data compression [4]. Data compression methods, in general, substitute one or more symbols with a code. A different approach One of these takes the place of one of these. A number of input symbols in a string. The term "floating point" is used to describe a The more time passes, the more The more complicated the information is encoded, the more For these purposes, a large number of bits are required. As well as This arithmetic coding produces one result. a less-than-one number and a number greater-than-one number or a value of zero. This number is one of a kind. decoded to generate series

METHODS

Text Data Compression

Compression is the process of converting a set of data into coded form with the aim of saving the need for storage space and time for data transmission[5] making it more efficient in storing it and shortening the data exchange time[6]. File compression is one of the important aspects in the development of information technology[7]. The compression process is carried out to reduce the consumption of memory resources, speed up the digital image transmission process[8]. The need for text data compression in the current cloud computing era is still quite high. Text data needs to be compressed as little as possible for easy transmission[9]. Data compression can be a solution to the problem above[10]. The technique of this compression is to replace the repetitive characters with a certain pattern so that the file can minimize its size[11]. Compression results depend on the selection of the desired compression quality[12].

Arithmetic Coding Method

A data compression approach was introduced with the aim of solving this problem[13]. This research is focused on the performance of the Arithmetic Coding algorithm[14] with implementation using VB 6.0. Arithmetic coding can be extended to compress correlated sources and has even better performance than channel coding for short to medium block length sources[15]. This algorithm performance analysis aims to determine the performance of this algorithm on files with .txt, .rtf, and .doc extensions. In this system there are compression and decompression stages. The compression stage aims to compress the file size, while the decompression stage aims to restore the file size to its original size[7]. The sequence of steps for the encoding process of this algorithm is as follows:

```
Set low = 0.0
Set high = 1.0
While (the input symbol still exists) do
  Take input symbol
  CR = high - low
  High = low + CR*high_range (symbol)
  Low = low + CR*low_range (symbol)
End While
- Print Low
```

In order for this software to run properly, there are several devices that must be met before executing this software. The tools needed to run this software include:

1. Hardware (Hardware)
 - a. Pentium IV microprocessor
 - b. The hard disk for the system to operate and as a data storage medium.
 - c. Minimum memory 1 Giga Byte
 - d. Super VGA Monitor
 - e. Mouse and Keyboard
 - f. Printer
2. Software (Software)
 - a. The operating system used must be at least Microsoft Windows XP or Microsoft Windows NT.
 - b. Visual Basic 6.0
 - c. Anti Virus to prevent possible system damage caused by viruses that enter either from diskettes or other input media.

RESULTS And DISCUSSION

Testing Data With VB 6.0

Algorithm design is a design that aims to provide an overview in coding the forms that have been designed. This Main Form display contains a menu display to receive input process mode and compression method as well as perform the work process of file compression. The main form display is as shown in Figure 1.

Figure 1: Main Menu Form

System Testing Techniques

After getting the results of the software display, the next stage the author conducts testing of the system. The system testing method that the author uses is a static method (static technique) where the test is divided into several stages, namely:

1. Setting Test Parameters
2. Setting Up the Test Kit
3. Performing Compression Test
4. Performing a Decompression Test

At this stage, testing is carried out on the ability of the system to compress text files and observe the

compression results produced. The first step is to select the compression mode and the results can be seen in Figure 2.

Figure 2: Compression Mode Selection Test Results

Next, the file to be compressed is added by pressing the Add button and the display form of this test is as shown in Figure 3.

Figure 3: Test Results Add Files

Next, input the destination file, and the form of the test results can be seen in Figure 4.

Figure 4: Testing Results of Destination File Name Input

In this stage, testing is carried out on the ability of the system to decompress files and observe the compression results produced. The first step is to select the decompression mode and the results can be seen in Figure 5.

Figure 5: Decompression Mode Selection Test Results

Next, add files to be decompressed by pressing the Load File button and the display form of this test is as shown in Figure 6.

Figure 6: Load File Test Results

Then input the destination file, and the form of the test results can be seen in Figure 7.

Figure 7: Test Results of the Destination File Name Input

Then press the Extract File button to decompress the Kompresi.arc file, then a message box will appear indicating the success or failure of the file decompression and the results can be seen in Figure 8.

Figure 8: File Decompression Test Results

CONCLUSION

Based on the results of the design and implementation of this TEXT file compression software using the arithmetic coding method, it can be concluded that: To compress a file, it can be done by simplifying the contents of files that have the same characteristics, so that in the storage process, the file size can be reduced. The arithmetic coding method can produce compression with a smaller size, up to half the size of the file before the compression process is carried out. The arithmetic coding method can compress files at a fairly high speed. The more contents of a file that have the same characteristics, then a file will be compressed to a size much smaller than the original size.

REFERENCES

- [1] R. Wulida Siam and M. Didik Rohmad Wahyudi, "Test the compression performance of Huffman trees one and two trees on text objects," *JNAALOKA*, 2020.
- [2] GI Yana and RK Hondro, "Implementation of the Huffman and Lz78 Algorithm for Data Compression," *J. Ris. computer.*, vol. 3, no. 6, 2016.
- [3] E. Suharyanto, "Implementation of Rivest Code 6 Cryptography-Based Data Security System, Vigenere Chipper and LZW Data Compression," *ESIT*. 2020.
- [4] D. Darwis, R. Prabowo, and N. Hotimah, "Combination of Gifshuffle, AES Encryption and Huffman Data Compression to Improve Data Security," *J. Teknol. inf. and Computer Science.*, vol. 5, no. 4, 2018.
- [5] AD Putri, D. Rachmawati, and H. Heriance, "Analysis and Implementation Playfair Cipher Cryptography Algorithm and Run Length Encoding Compression Algorithm for Security and

- Text Data Compression,” *Talent. conf. Ser. science. Technol.*, vol.1, no. 1, 2018.
- [6] CT Utari, "Implementation of Run Length Encoding Algorithm for Image File Compression and Decompression Application Design," *J. TIMES*, vol. V, no. 2, 2016.
 - [7] HT Sihotang, "Design and Implementation of Arithmetic Coding Algorithm for Video and Audio Data Compression Applications," *J. Mantik Penusa*, vol. 2, no. 1, 2018.
 - [8] AM Faza, C. Slamet, and D. Nursantika, "Performance Analysis of Digital Image Compression with DWT, DCT and Hybrid Comparison (DWT-DCT)," *J. Online Information.*, vol. 1, no. 1, 2016.
 - [9] SM Sunaryo, L. Chrisantyo, and Y. Lukito, "ANALYSIS OF MTF, MTF-1 AND MTF-2 ALGORITHM ON BURROWS WHEELER COMPRESSION ALGORITHM," *J. Apply. Technol. inf.*, vol. 3, no. 1, 2019.
 - [10] C. ALIWARGA, AA PRAMUDITA, and MA KARTAWIDJAJA, "ECG Signal Compression using Parameter Extraction Technique," *ELKOMIKA J. Tek. Electrical Energy. Tech. Telecom. Tech. Electron.*, vol. 7, no. 2, 2019.
 - [11] MA Latif, SD Nasution, and P. Pristiwanto, "Comparative Analysis of Rice Codes Algorithm with Goldbach Codes Algorithm in Text File Compression Using Exponential Method," *Maj. science. CORE (Information and Technol. Scientific)*, vol. 13, no. 1, 2018.
 - [12] R. Krasnala, A. Budimansyah, and UT Lenggana, "Image Compression By Combining Discrete Cosine Transform (DCT) Method and Huffman Algorithm," *J. Online Information.*, vol. 2, no. 1, 2017.
 - [13] GA Pradnyana and IBP Suarma Putra, "DIGITAL DATA FILE SECURITY WITH COMBINATION OF TRIPLE TRANSPOSITION CIPHER ALGORITHM AND HUFFMAN METHOD," *J. Educator. Technol. and Vocational.*, vol. 15, no. 1, 2018.
 - [14] NA Khairi, AB Jambek, and RC Ismail, "Performance evaluation of arithmetic coding data compression for internet of things applications," *Indonesia. J. Electr. eng. Comput. science.*, vol. 13, no. 2, 2019.
 - [15] J. Zhou, Y. Fu, Y. Yang, and ATS Ho, "Distributed video coding using interval overlapped arithmetic coding," *Signal Processing. Image Commun.*, vol. 76, 2019.