



PROJECT REPORT
LZ77 AND SHANNON-FANO TEXT COMPRESSION
ALGORITHM COMPARISON

Rudi Christanto Yuwono
10.02.0001
2013/2014

FACULTY OF COMPUTER SCIENCE
SOEGIJAPRANATA CATHOLIC UNIVERSITY
Jl. Pawiyatan Luhur IV/1, Bendan Duwur, SEMARANG 50234
Telp. 024-8441555 (hunting) Web: <http://www.unika.ac.id>
<http://ikomunika.web.id/>

i

PERPUSTAKAAN	
Universitas Katolik Soegijapranata	
No. Inv.	257 / S / IK / c.1
Tanggal	19 Agustus 2014
Paraf	

PPROVAL AND RATIFICATION PAGE

PROJECT REPORT

LZ77 and Shannon-Fano Text Compression Algorithm Comparison

by

10.02.0001 – Rudi Christanto Yuwono

This project report has been approved and ratified by the Dean of Faculty of Computer
Science and Supervisor on 18 July 2014

With approval,

Examiners,


Hironimus Leong, S.Kom., M.Kom
NPP : 058.1.2007.273

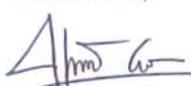
Supervisor,


Rosita Herawati, ST., MIT
NPP : 058.1.2004.263

Examiners,


Suyanto Edward Antonius, Ir., M.Sc
NPP : 058.1.1992.116

Examiners,


Shinta Estri Wahyuningrum, S.Si, M.Cs
NPP : 058.1.2007.272

Dean of Faculty of Computer Science,



Hironimus Leong, S.Kom., M.Kom
NPP : 058.1.2007.273

STATEMENT OF ORIGINALITY

I, the undersigned:

Name : Rudi Christanto Yuwono

ID : 10.02.0001

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.

Semarang, 18 July 2014



Rudi Christanto Yuwono

10.02.0001

ABSTRACT

There are many ways of text compression, but they are all grouped by two large groups: lossless and lossy compression, and from that alone, the lossless method can still be divided into two types: Statistical Compression and Dictionary-based compression. This project will examine the effectiveness of these two methods, which one is more effective for compressing text files. Shannon – Fano Algorithm is used to represent Statistical Compression and Lempel-Ziv 1977(LZ77) algorithm is used to represent dictionary-based compression.

The final result is, the program is able to compress using the two aforementioned algorithms and generate the output file. But the file generation cannot work in the graphical mode, only in text mode. And the LZ77-compressed file always has smaller file size than the Shannon-Fano-compressed one.

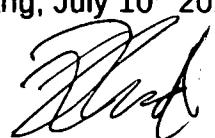
FOREWORD

First of all, I lift my praises and gratitude to the Almighty God, Lord Jesus Christ, and Holy spirit, whom have guided me and empowered me in working on this project to the very end. Also, I want to express my gratitude to these parties:

- First and foremost, for God, Lord Jesus, and Holy Spirit who empowered me to finish this project, passing through every obstacles and many unexpected things.
- Both of my parents, who have supported me to push forward no matter what, and cheered me up whenever I'm down.
- Every single lecturer in Faculty of Computer Science, for all of the lectures that enabled me to know what I'm doing. Your teachings are very meaningful.
- All of my friends from Ikorn 10, Nathanael Youth Community, GCT Kaskus, and Gensokyo.co.id Community. We have fought hard together for these 4 years, we put up a good fight, and finally we finish together, through many tears and joy.
- Google, Wikipedia, and CiteseerX. A student's best friend. Their role is completely indispensable in this project.
- Facebook, Youtube, and 9GAG. The ultimate procrastination station.
- The doctor whom I visit whenever my health drops down. Sorry for coming too often.
- My notebook PC. Of course, without this thing, this whole project can never ever be done.

I fully understand that the project still has many flaws and very far from the word "perfect". I look forward for suggestions and criticisms.

Semarang, July 10th 2014



Rudi Christanto Yuwono

10.02.0001

TABLE OF CONTENTS

APPROVAL AND RATIFICATION PAGE.....	ii
STATEMENT OF ORIGINALITY.....	iii
ABSTRACT.....	iv
FOREWORD.....	v
TABLE OF CONTENTS.....	vi
TABLE OF FIGURES.....	viii
TABLE OF TABLES.....	ix
CHAPTER I.....	1
INTRODUCTION.....	1
CHAPTER II.....	3
LITERATURE STUDY.....	3
2.1 Algorithm.....	3
2.1.1 LZ77 Algorithm.....	3
2.1.2 Shannon – Fano algorithm.....	3
2.1.3 The Sliding Window.....	4
2.1.4 Shannon–Fano's Frequency Table.....	4
CHAPTER III.....	6
PLANNING.....	6
3.1 Research Methodologies.....	6
3.2 Project Management.....	7
CHAPTER IV.....	8
ANALYSIS AND DESIGN.....	8
4.1 Analysis.....	8
4.1.1 Use case diagram.....	8

4.1.2 Class diagram.....	9
4.2.2 Flow Chart.....	10
CHAPTER V.....	11
IMPLEMENTATION AND TESTING.....	11
5.1 Implementation.....	11
5.2 Testing.....	18
5.3 Main Interface Window.....	23
CHAPTER VI.....	35
CONCLUSION AND FURTHER RESEARCH.....	35
6.1 Conclusions.....	35
6.2 Further Research.....	35
References.....	36

TABLE OF FIGURES

Figure 2.1.4.1 Frequencies Table.....	4
Figure 2.1.4.2 The Coding Process.....	5
Figure 2.1.4.3 The table with the code.....	5
Figure 4.1.1 Use case diagram.....	8
Figure 4.1.2 Class Diagram.....	9
<i>Figure 4.1.3 Flow Chart.....</i>	10
Figure 5.3.1 The main interface window.....	23
Figure 5.3.2 The file selector.....	24
Figure 5.3.3 Showing the contents of test3.txt.....	25
Figure 5.3.4 LZ77 compressed form.....	26
Figure 5.3.5 Shannon – Fano compressed form.....	27
Figure 5.3.6 LZ77 compression in text mode, with proper output file generation and its file size.....	28
Figure 5.3.7 Shannon - Fano compression in text mode, with proper output file generation and its file size.....	28
Figure 5.3.8 The unchanged compressed file size.....	29
Figure 5.3.9 Testing with 2-characters length.....	30
Figure 5.3.10 Testing with 3-characters length.....	30
Figure 5.3.11 Testing with 4-characters length.....	31
Figure 5.3.12 Testing with a file without a possible match (abcdefghijkl), generates 8 bytes from 12 bytes file.....	32
Figure 5.3.13 Testing with a file with a few matches (aaabbcaaaabb), generates 4 bytes from 12 bytes file.....	32
Figure 5.3.14 Testing with a file with a definite match (aaaaaaaaaaa), generates 3 bytes from 12 bytes.....	33

TABLE OF TABLES

Table 3.2 Time schedule table.....	7
Table 5.2 Test result table.....	22