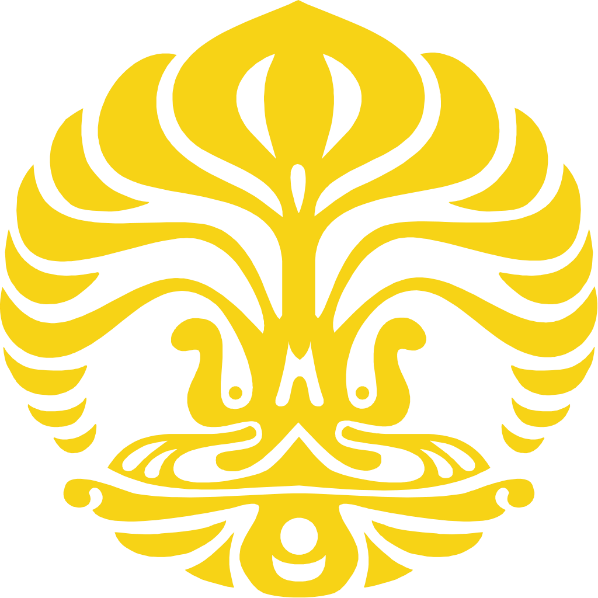
SOFTWARE ENGINEERING PROPOSAL

LOSSLESS DATA FILE COMPRESSION SOFTWARE



**Aria Lesmana (1506732690) (41)**

**Johanes Gunawan (1506732690) (34)**

**Wildan Haykal (1506725413) (35)**

**Chapter 1**

* 1. **General Purpose**

There are two major categories of compression algorithms: lossy and lossless. Lossy compression algorithms involve the reduction of a file’s size usually by removing small details that require a large amount of data to store at full fidelity. In lossy compression, it is impossible to restore the original file due to the removal of essential data. Lossy compression is most commonly used to store image and audio data, and while it can achieve very high compression ratios through data removal, it is not covered in this article. Lossless data compression is the size reduction of a file, such that a decompression function can restore the original file exactly with no loss of data. Lossless data compression is used ubiquitously in computing, from saving space on your personal computer to sending data over the web, communicating over a secure shell, or viewing a PNG or GIF image.

The basic principle that lossless compression algorithms work on is that any non-random file will contain duplicated information that can be condensed using statistical modeling techniques that determine the probability of a character or phrase appearing. These statistical models can then be used to generate codes for specific characters or phrases based on their probability of occurring, and assigning the shortest codes to the most common data. Such techniques include entropy encoding, run-length encoding, and compression using a dictionary. Using these techniques and others, an 8-bit character or a string of such characters could be represented with just a few bits resulting in a large amount of redundant data being removed.

The ZIP format and other DEFLATE-based formats were king up until the mid 1990s when new and improved formats began to emerge. In 1993, Eugene Roshal released his archiver known as WinRAR which utilizes the proprietary RAR format. The latest version of RAR uses a combination of the PPM and LZSS algorithms, but not much is known about earlier implementations. RAR has become a standard format for sharing files over the Internet, specifically in the distribution of pirated media. An open-source implementation of the Burrows-Wheeler Transform called bzip2 was introduced in 1996 and rapidly grew in popularity on the UNIX platform against the DEFLATE-based gzip format. Another open-source compression program was released in 1999 as the 7-Zip or .7z format. 7-Zip could be the first format to challenge the dominance of ZIP and RAR due to its generally high compression ratio and the format's modularity and openness. This format is not limited to using one compression algorithm, but can instead choose between bzip2, LZMA, LZMA2, and PPMd algorithms among others.

Finally, on the bleeding edge of archival software are the PAQ\* formats. The first PAQ format was released by Matt Mahoney in 2002, called PAQ1. PAQ substantially improves on the PPM algorithm by using a technique known as context mixing which combines two or more statistical models to generate a better prediction of the next symbol than either of the models on their own.

Nowadays, the security of files and efficient usage of space in computers become more essential and significantly needed, because of advancement in computer tech, data files are getting larger and data sharing over long distance is in popular demand.   
  
Lossless data compression is the size reduction of a file, such that a decompression function can restore the original file exactly with no loss of data. Lossless data compression is used ubiquitously in computing, from saving space on personal computers, sending data over the web, communicating over a secure shell, or viewing an image file.

By reducing the total data size, file sharing using compressed file can be done faster than using uncompressed file, compression can also secure the files using encryption. The main purpose of this software is to store a group of files into a package file which is smaller in size than the total size of the group of files through encryption encoding methods.

* 1. **Function**
* Archiving and compressing multiple files into one package file
* Reducing storage burden by reducing packaged file size through compression methods
* Increasing security level of data
  1. **Tools**

- C++ programming language

- GNU Compiler, GCC

- Github

- TASM/Emu8086

* 1. **Hardware**
* PC

The target of this project is to archive and compress one or multiple files, so the main tool for this project is a PC.

* Storage Unit

The main function of this software is to maximize efficiency of storage unit by compressing large file(s) into smaller size archive file.

* 1. **Risk Analysis**
* Compressed file corruption

In case of compressing file, there is a risk of failure after decompression which will leave the file corrupted or unusable. This risk can be solved by having a backup file but it is in contrast to the function of reducing storage burden so we need to make sure that the file will not be corrupt.

* Inefficient compression time

One of the function of this software is to make file sharing become more efficient and faster. Which means that using a lot of time in compressing or decompressing data is one of the problem that obstruct the function

* Bug may stop compression midway

Another problem that needed to be solved is bug which can come in a lot of way. One of them is the software will stop responding during a file compression process. This can result in a corrupted file.

Chapter 2

2.1 Project Integration Management

Project integration management aiming to coordinate all knowledge in the area of project management through the project process, includes the processes and activities needed to identify, define, combine, unify, and coordinate the various things that needed for completion. This also ensures that all of the elements will be successfully completing the project on time. The main activities in management integration are

* Project Plan Development

For plan development, this file compression software will provide lossless data compression and simple interface to use and execute the compression or decompression process

* Project Plan Execution

To complete those plans, there are several executions that have been made just like analyze what will the projects do. In details the execution is written in time management part together with the tasks. In order to make the application, it needs some requirement,

2.2 Project Scope Management

The function of project scope management is to make sure this project includes all the work relevant to achieving the project’s objectives. It is a description of all the work that must be done on a project. It includes all of the project deliverables. There are several points for this scope which is:

* Scope Planning

The whole planning is written along this documentation. Goals, objectives, and any other requirements are already mentioned above. Besides that, project success is determined by its completion, usefulness and also : in increase of revenue in savings of costs.

* Scope Definition

The project is called lossless data compression software. The goal of this project is to make a more efficient compression software. This software is built to be compatible with windows and linux.

* Scope Verification

Scope verification is the process of formalizing acceptance of the project scope by the stakeholders. It requires reviewing work products and results to ensure that all were completed correctly and satisfactorily. During the course of this project, several documents and programs have to be produced and delivered at the end of project :

* SPMP (Software Project Management Plan)
* Design Diagram
* Test Plan
* User’s Manual

2.3 Project Time Management

This kind of management is very important because time has the least flexibility so we have to manage very well. Based on this timeline, project-planning and analysis are written in this report.

2.4 Project Cost Management

Project Cost Management is a series of activities for estimating, allocating, and controlling costs within the project. It allows determining and approving budget for the project and controlling spending.

* Cost Estimating, Budgeting and control

There’s no cost of manufacture of this project, because we only need to install software to develop this.

* Resources Planning

Because we don’t have need any cost for this project, maybe we just spend for transportation costs in the need of meet with others.

2.5 Project Quality Management

Quality management is the process for ensuring that all project activities necessary to design, plan and implement a project are effective and efficient with respect to the purpose of the objective and its performance.

* Quality Planning

This project is about compressing. The quality that we have planned are :

* Easy to use
* Can compress file efficiently
* Simple manual
* Simple algorithm explanation

2.6 Project Human Resources Management

Project Human Resource Management process involves identifying and documenting project roles and responsibilities. The processes required to make the most effective use of the people involved with the project. According to the PMI, this process divided by 4 main points :

* Develop human resource plan
* Required skills
* Organizational relationships
* Creating a staffing plan

2.7 Project Communication Management

- Communications Plan

For this planning we thought that we won’t meet any difficulties, considering the communication will be easy to do because we often meet at UI. For support, we use Line application to do online sharing or anything urgent and need to be solved together.

2.8 Project Risk Management

Project risk management is the art and science of identifying, analyzing, and handling risks through the process of the project with the main purpose of meeting project objectives. Risk management has a positive impact on the selection of the project, defining the scope of the project, develop a realistic schedule, estimated costs to be incurred, and the most important is to handle errors from the application.

2.8.1 Risk Identification

The risks of this project are focused to the weakness and error-features that may be happen next. Five categories of risk are identified :

* Programming skill, medium or higher is needed
* Low design
* Time shortage high
* Satisfaction

2.8.2 Risk Quantification

If we can’t handle and solve the risks, so it can make our project goals fails, and the clients will be disappointed to not get the advantages from the features.

2.8.3 Risk Response Development

To minimize the risks, the most important thing is we need to work hard in programming and manage time efficiently.

2.8.4 Risk Response Control

For each points of the risk, the response that we can do is planned like this :

* Learning programming skill
* Learning the optimal algorithm of compressing
* Learning to design

2.9 Project Procurement Management

This Procurement Management Plan sets the procurement framework for Logic Learning Games for Children application. It will serve as a guide for managing the procurements throughout the life of the project and will be updated as acquisition needs change. This plan identifies and defines the items to be procured, the types of contracts to be used in support of this project, the contract approval process, and decision criteria. The importance of coordinating procurement activities, establishing firm contract deliverables, and metrics in measuring procurement activities is also included.

2.9.1 Procurement Planning

The Project Leader will provide oversight and management in concert with appropriate agency procurement and management staff for all procurement activities under Logic Learning Games for Children. The Project Leader will work with the project team to identify all items to be procured for the successful completion of the project.