

ABSTRACT

The Domain “Sun Zip” lets you reduce the overall number of bits and bytes in a file so it can be transmitted faster over slower Internet connections, or take up less space on a disk. Domain Sun Zip is a System Based Software. The user need not depend on third party software’s like winzip, winrar, Stuff etc.

The main algorithm is:

- Huffman algorithm

Algorithms for Huffman Compression/Decompression

Huffman coding is an entropy encoding algorithm used for lossless data compression. The term refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol. It was developed by David A. Huffman while he was a Ph.D. student at MIT, and published in the 1952 paper "A Method for the Construction of Minimum-Redundancy Codes".

Huffman coding uses a specific method for choosing the representation for each symbol, resulting in a prefix code (sometimes called "prefix-free codes") (that is, the bit string representing some particular symbol is never a prefix of the bit string representing any other symbol) that expresses the most common characters using shorter strings of bits than are used for less common source symbols. Huffman was able to design the most efficient compression method *of this type*: no other mapping of individual source symbols to unique strings of bits will

produce a smaller average output size when the actual symbol frequencies agree with those used to create the code. A method was later found to do this in linear time if input probabilities (also known as *weights*) are sorted.

For a set of symbols with a uniform probability distribution and a number of members which is a power of two, Huffman coding is equivalent to simple binary block encoding, e.g., ASCII coding. Huffman coding is such a widespread method for creating prefix codes that the term "Huffman code" is widely used as a synonym for "prefix code" even when such a code is not produced by Huffman's algorithm.

Although Huffman coding is optimal for a symbol-by-symbol coding (i.e. a stream of unrelated symbols) with a known input probability distribution, its optimality can sometimes accidentally be over-stated. For example, arithmetic coding and LZW coding often have better compression capability. Both these methods can combine an arbitrary number of symbols for more efficient coding, and generally adapt to the actual input statistics, the latter of which is useful when input probabilities are not precisely known or vary significantly within the stream. In general, improvements arise from input symbols being related (cat is more common than cta).

1. OVERVIEW OF THE PROJECT

1.1 INTRODUCTION

The Domain “File Compression” lets you reduce the overall number of bits and bytes in a file so it can be transmitted faster over slower Internet connections, or take up less space on a disk. Domain File compression is a System Based Software. The software will be done using Core Java. It can use in the System as a utility. The type of compression we will use here is called **lossless compression**. The user need not depend on third party software’s like winzip, winrar, Stuff etc. the software can be used to compress files and they can be decompressed when the need arises. For implementing this Software we want to use algorithms

The main algorithm is:

- Huffman algorithm

The Domain File Compression mainly include 7 modules

- **Compress A File Or Folder**
- **De-Compress the file or folder**
- **View files in the compressed file**
- **Facility to set icon**
- **Facility to set your own extension**

1. Compress file or folder

This module helps us to compress a file or folder. The compressed file will have a extension that has been given at the development time. We can send the compressed file over the internet so that users having this software can decompress it.

2. Decompress a file or folder

This is the reverse process of file compression. Here we can decompress the compressed file and get the original file.

4. View files in the compressed file

Here we can view the list of files inside our compressed file. We can view the files before decompressing and decide to decompress or not.

5. Set icon and extension

This is additional feature in our project. We can set our own extension to the compressed file. More than that we can specify the style of icon for the compressed file. Users will also be given a option to change the icon as per their preference.

1.2 APPLICATION AREAS

The application areas of file compression are

- File storage
- Distributed systems.

2

SYSTEM STUDY AND ANALYSIS

System analysis is a process of gathering and interpreting facts, diagnosing problems and the information to recommend improvements on the system. It is a problem solving activity that requires intensive communication between the system users and system developers. System analysis or study is an important phase of any system development process. The system is studied to the minutest detail and analyzed. The system analyst plays the role of the interrogator and dwells deep into the working of the present system. The system is viewed as a whole and the input to the system are identified. The outputs from the organizations are traced to the various processes. System analysis is concerned with becoming aware of the problem, identifying the relevant and decisional variables, analyzing and synthesizing the various factors and determining an optimal or at least a satisfactory solution or program of action.

A detailed study of the process must be made by various techniques like interviews, questionnaires etc. The data collected by these sources must be scrutinized to arrive to a conclusion. The conclusion is an understanding of how the system functions. This system is called the existing system. Now the existing system is subjected to close study and problem areas are identified. The designer now functions as a problem solver and tries to sort out the difficulties that the enterprise faces. The solutions are given as proposals. The proposal is then weighed with the existing system analytically and the best one is selected. The proposal is presented to the user for an endorsement by the user. The proposal is reviewed on user request

and suitable changes are made. This is loop that ends as soon as the user is satisfied with proposal.

Preliminary study is the process of gathering and interpreting facts, using the information for further studies on the system. Preliminary study is problem solving activity that requires intensive communication between the system users and system developers. It does various feasibility studies. In these studies a rough figure of the system activities can be obtained, from which the decision about the strategies to be followed for effective system study and analysis can be taken.

Here in the project SunZip, a detailed study of existing system is carried along with all the steps in system analysis. An idea for creating a better project was carried and the next steps were followed.

2.1 FEASIBILITY STUDY

An important outcome of the preliminary investigation is the determination that the system requested is feasible. Feasibility study is carried out to select the best system that meets the performance requirements.

Feasibility study is both necessary and prudent to evaluate the feasibility of the project at the earliest possible time. It involves preliminary investigation of the project and examines whether the designed system will be useful to the organization. Months or years of effort, thousand for millions of money and untold professional embarrassment can be averted if an in-conceived system is recognized early in the definition phase.

The different types of feasibility are: Technical feasibility, Operational feasibility, Economical feasibility.

1) Technical feasibility

Technical Feasibility deals with the hardware as well as software requirements. Technology is not a constraint to type system development. We have to find out whether the necessary technology, the proposed equipments have the capacity to hold the data, which is used in the project, should be checked to carryout this technical feasibility.

The technical feasibility issues usually raised during the feasibility stage of investigation includes these

- This software is running in windows 2000 Operating System, which can be easily installed.
- The hardware required is Pentium based server.
- The system can be expanded.

2) Operational feasibility

This feasibility test asks if the system will work when it is developed and installed.

Operational feasibility in this project:

- The proposed system offers greater level of user-friendliness.

- The proposed system produces best results and gives high performance. It can be implemented easily .So this project is operationally feasible.

3) Economical feasibility

Economical Feasibility deals about the economical impact faced by the organization to implement a new system. Financial benefits must equal or exceed the costs. The cost of conducting a full system, including software and hardware cost for the class of application being considered should be evaluated.

Economic Feasibility in this project:

- The cost to conduct a full system investigation is possible.
- There is no additional manpower requirement.
- There is no additional cost involved in maintaining the proposed system.

EXISTING SYSTEM

Existing system refers to the system that is being followed till now. The main disadvantage of this system is that the users depend on third party software's like winzip, winrar, Stuff etc.

The existing system requires more computational time, more manual calculations, and the complexity involved in Selection of features is high. The other disadvantages are lack of security of data, Deficiency of Data accuracy, Time consuming etc.

To avoid all these limitations and make the working more accurately the system needs to be computerized.

Draw backs of existing system.

- Lack of security of data.
- Deficiency of Data accuracy
- Time consuming.
- The users depend on third party software's like winzip, winrar, Stuff etc.

To avoid all these limitations and make the working more accurately the system needs to be computerized.

PROPOSED SYSTEM

The aim of proposed system is to develop a system of improved facilities. The proposed system can overcome all the limitations of the existing system. The system provides data accuracy and save disc space. The existing system has several disadvantages and many more difficulties to work well. The proposed system tries to eliminate or reduce these difficulties up to some extent. The proposed system is file/folder compression or decompression based on the Huffman algorithm and GZip algorithm. The proposed system will help the user to consume time. The proposed system helps the user to work user friendly and he can easily do the file compression process without time lagging. The system is very simple in design and to implement. The system requires very low system resources and the system will work in almost all configurations. It has got following features Ensure data accuracy, minimize manual data entry, minimum time needed for the various processing, greater efficiency, better service.

Advantages of Proposed System

The system is very simple in design and to implement. The system requires very low system resources and the system will work in almost all configurations. It has got following features

- Ensure data accuracy and Save disk space
- Minimum time needed for the file compression
- Greater efficiency and Better Service
- Protection from virus and Easy to send via E-mail

- Maximum Compression rate is 2 GB.
- The user need not depend on third party software's like winzip, winrar, Stuff etc.

SYSTEM DESIGN

System Design is the most creative and challenging phase in the system life cycle. Design is the first step into the development phase for any engineered product or system. Design is a creative process. A good design is the key to effective system. System design is a solution *how to approach* the creation of a new system. System design transforms a logic representation of what is required to do into the physical specification. The specification is converted into physical reality during development.

LOGICAL DESIGN

The logical flow of a system and define the boundaries of a system. It includes the following steps:

- Reviews the current physical system – its data flows, file content, volumes, frequencies etc.
- Prepares output specifications – that is, determines the format, content and frequency of reports.
- Prepares input specifications – format, content and most of the input functions.
- Prepares edit, security and control specifications.
- Specifies the implementation plan.
- Prepares a logical design walk through of the information flow, output, input, controls and implementation plan.
- Reviews benefits, costs, target dates and system constraints.

PHYSICAL DESIGN

Physical system produces the working systems by define the design specifications that tell the programmers exactly what the candidate system must do. It includes the following steps.

- Design the physical system.
- Specify input and output media.
- Design the database and specify backup procedures.
- Design physical information flow through the system and a physical design walk through.
- Plan system implementation.
- Prepare a conversion schedule and target date.
- Determine training procedures, courses and timetable.
- Devise a test and implementation plan and specify any new hardware/software.
- Update benefits , costs , conversion date and system constraints

Design/Specification activities

- Concept formulation.
- Problem understanding.
- High level requirements proposals.
- Feasibility study.

- Requirements engineering.
- Architectural design.

INPUT DESIGN

Input Design deals with what data should be given as input, how the data should be arranged or code, the dialog to guide the operating personnel in providing input, methods for preparing input validations and steps to follow when error occur. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

In this project, the input design consists of a log in screen, tab for compression/ decompression, source and destination browsing button, a menu list for selecting the algorithm, Compress/Decompress option, compress/decompress button.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. The objective of output design is to convey information about past activities, current status or projections of the future, signal important events, opportunities, problems, or warnings, trigger an action, confirm an action etc. Efficient, intelligible output design should improve the system's relationship with the user and helps in decisions making. In output design the emphasis is on displaying the output on a CRT screen in a predefined format. The primary consideration in design of output is the information requirement and objectives of the end users. The major formation of the output is to convey the information and so its layout and design need a careful consideration.

There is a output display screen for showing the compressed/ decompressed file or folder details(Original file size, Compressed/Decompressed file size, Distinct characters)

DATABASE DESIGN

A database is an organized mechanism that has the capability of storing information through which a user can retrieve stored information in an effective and efficient manner. The data is the purpose of any database and must be protected.

The database design is a two level process. In the first step, user requirements are gathered together and a database is designed which will meet these requirements as clearly as possible. This step is called Information Level Design and it is taken independent of any individual Database Management System (DBMS).

In the second step, this Information level design is transferred into a design for the specific DBMS that will be used to implement the system in question. This step is called Physical Level Design, concerned with the characteristics of the specific DBMS that will be used. A database design runs parallel with the system design. The organization of the data in the database is aimed to achieve the following two major objectives.

- Data Integrity
- Data independence

SOFTWARE DESCRIPTION

This project is implemented using Java. Java goes back to 1991 when a group of sun engineers led by James Gosling , wanted to design a small computer language that could be used for consumer devices and named it as Green Project. Their idea was to develop a portable language that could generate intermediate code for virtual machines. This intermediate code then can be used on any machines that has the correct interpreter.

Java is a programming language that lets us to do almost anything we can do with traditional programming language for distributed applications. It is platform in-dependent and having a lot of networking features included within it. A java program can run equally well on any architecture that has a java interpreter.

2.2.1 FEATURES OF JAVA

1) Encapsulation

Data Encapsulation is one of the most sticking features of OOP's. Encapsulation is the wrapping up of data and function into single unit called class. The wrapped defines the behaviour and protects the code and data from being arbitrarily accessed by the outside world and only those function which are wrapped in the class can access it. This type of insulation of data from direct access by the program is called data hiding.

2) Inheritance

Inheritance is the process by which objects a class can acquire the properties of objects of another class i.e. In OOPs the concept of inheritance provides idea of reusability providing the means of adding additional features to an existing class without modifying it. This is possible by deriving a new class from the existing one thus the newly created class will have the combined features of both the parent and the child classes.

3) Object –Oriented

Almost everything in java is a class, a method or an object. Only the most basic primitive operative and data types are at a sub-class level.

4) Data Abstraction

Data Abstraction is an act of representing essential features without including the background details and explanation.

5) Platform Independent

Java programs are compiled with a byte code format that can be read and run by interpreters on many platforms including Windows 95, Windows NT and later.

6) Multi-Thread

Java is inherently multi-threaded. A single java program can make many different things processing independently and continuously.

7) High Performance

Java can be compiled on the fly with a Just-in-time compiler (JIT) to code that rivals C++ in speed.

8) Safe

Java code can be executed in an environment that prohibits it from viruses, deleting or modifying files or otherwise performing data destroying and computer crashing operation.

9) Simple

Java has the bare bones functionally needed to implement its rich feature set.

2.2.2 Components

Java has several in-built components:

Javac : Compiler for java programs that could generate byte codes

Java : Interpreter to read and execute java byte codes.

Javap : To disassemble and debug the java bytecodes.

Javadoc: Document generator.

Javah : To write and link native codes with java programs.

SYSTEM TESTING AND IMPLEMENTATIONS

Testing is a process of executing a program with the interest of finding an error. A good test is one that has high probability of finding the yet undiscovered error. Testing should systematically uncover different classes of errors in a minimum amount of time with a minimum amount of efforts.

Two classes of inputs are provided to test the process

1. A software configuration that includes a software requirement specification, a design specification and source code.
2. A software configuration that includes a test plan and procedure, any testing tool and test cases and their expected results.

Testing is divided into several distinct operations:

1. Unit Testing

Unit test comprises of a set tests performed by an individual program prior to the integration of the unit into large system. A program unit is usually the smallest free functioning part of the whole system. Module unit testing should be as exhaustive as possible to ensure that each representation handled by each module has been tested. All the units that makeup the system must be tested independently to ensure that they work as required.

During unit testing some errors were raised and all of them were rectified and handled well. The result was quiet satisfactory and it worked well.

2. Integration Testing

Integration testing is a system technique for constructing the program structure while at the same time conducting tests to uncover errors associated with interfacing. The objective is to take unit tested modules and build a program structure that has been dictated by design. Bottom-up integration is the traditional strategy used to integrate the components of a software system into functioning whole. Bottom-up integration consists of unit test followed by testing of the entire system. A sub-system consists of several modules that communicated with other defined interface.

The system was done the integration testing. All the modules were tested for their compatibility with other modules .They test was almost successful. All the modules coexisted very well, with almost no bugs. All the modules were encapsulated very well so as to not hamper the execution of other modules.

3. Validation Testing

After validation testing, software is completely assembled as a package, interfacing errors that have been uncovered and corrected and the final series of software test; the validation test begins. Steps taken during software design and testing can greatly improve the probability of successful integration in the larger system. System testing is actually a series of different tests whose primary purpose is to fully exercise the compute –based system.

4. Recovery Testing

It is a system that forces the software to fail in a variety of ways and verifies that the recovery is properly performed.

5. Security Testing

It attempts to verify that protection mechanisms built into a system will in fact protect it from improper penetration. The system's security must of course be tested from in vulnerability form frontal attack.

6. Stress Testing

Stress tools are designed to confront programs with abnormal situations. Stress testing executes a system in a manner that demands resources in abnormal quantity and volume.

7. Black Box Testing

Black box testing is done to find out the following information as shown in below:

1. Incorrect or missing functions.
2. Interface errors.
3. Errors or database access.

4. Performance error.
5. Termination error.

The mentioned testing is carried out successfully for this application according to the user's requirement specification.

8. Test Data Output

After preparing test data, the system under study is tested using the test data. While testing the system using test data, errors are again uncovered and corrected by using above testing and corrections are also noted for future use.