**Project outline**

**Name:jionglin Tao**

**Student ID:120109498**

1. **Research Area**

The research area of this project is lossless compression algorithm, compression using statistical redundancy of the data allows full recovery of the original data without any distortion. And this project uses the bloom filter which is a data structure to detect whether an element is a member of a collection to implement lossless compression.

1. **Objectives**

Implement a new lossless compression and decompression algorithm, and evaluate the algorithms against one or two existing algorithms.

1. **Research Questions**

• How to eliminate false positives?

• How to improve the compression ratio?

• How to optimize the number and type of hash functions?

1. **Deliverables**

▪An evaluation of new lossless compression algorithm

▪An analysis of impact of different number of hash functions

▪An analysis of impact of different type of hash functions

▪A report of comparison of the new algorithm with the existing algorithms

1. **Timetable**

|  |  |  |
| --- | --- | --- |
|  | **Description of Work** | **Detail** |
| **Phase One** | Implement the new lossless compression algorithm | Programming language:Java  Main data structures: Bloom filter,Witness.  Main classes:   * Hash: To generate the hash value. * Encoder: Implement bloom filter with witness to compress the input * Decoder: To decompress the encoded data.   Main Interface:   * HashFunction: To provide an abstract method to implement different hash functions. |
| **Phase Two** | Testing and optimizing | Test the code, and calculate the compression ratio,use the different seed.  To get the optimized number and type of hash functions. |
| **Phase Three** | Compare with existing algorithms | Compare the compression ratio, running time,etc. |