Project Proposal

Intro:

Background:

Nowadays,Everyone saves most of their documents in a digital way. This is good because it means that they don't have to carry physical documents with them and this means that they can view their documents anywhere in the world. But, as we all know digital files will take up a lot of memory space.

For example, in the 2017 Panama Papers leak, 13.4 million secret documents were leaked to the public. These documents took up to 2 terabytes of data therefore we believe that compressing these file could have saved up a lot of memory space.

So, For our analysis of algorithms final project, we decided to compare the results between 2 lossless compression algorithms which can be used to compress text files.

https://www.forbes.com/sites/niallmccarthy/2017/11/06/the-scale-of-the-paradise-papers-leak-infographic/#5a30df956a76

Problem Description:

To compare the time complexity,space complexity, and compression ratios between two lossless compression algorithms, LZ-77 and Lempel-Ziv-Welch when used to compress documents(text files in this case).

Related Work(s):

Other programs that try to solve this problem include Winrar,7zip,winzip and more.These programs compress the files by removing redundant data from the compressed file and re-add the redundant data when you decompress the file.

Proposed Method(s):

Algorithms Used:

- **LZ-77:** Replaces repeated occurrences of data with references to a single copy of that data existing earlier. A match is encoded by a pair of numbers called a *length-distance pair*, which is equivalent to the statement "each of the next *length* characters is equal to the characters exactly *distance* characters behind it in the uncompressed stream".

- **Lempel-Ziv-Welch Encoding**: Create a dictionary (a table) of strings used during the communication session.The idea is to parse the sequence into distinct phrases, because when a node is inserted, the code for the current piece becomes the parent node combined with the new character.

Why we chose these algorithms ?

We chose to use these algorithms because Lempel-Ziv-Welch was regarded as one of the best lossless algorithms and LZ-77 is one of the most popular lossless algorithm.

Our Activities Step by Step:

- Research about LZ-77 encoding and Lempel-Ziv-Welch understand how these algorithms work.
- Try to code programs implementing these algorithms.
- Review the code and analyze for bugs.
- Compare the compression and decompression results from both programs.
- Explain algorithms and compression results in report