Summary of 2013 Project

2013 project was basically a simple research about text compression and the analysis of the first of Pessoa's algorithms. So, in the research we found that the first attempt prefix free codes was by Shannon-Fano in 1948. Later, in 1951 Fano challenged his students to improve his algorithm. So, in 1952, Huffman Coding appeared, from one of the students of Fano: David Huffman. (reference needed)

So with this historical introduction, we proceeded to explain Huffman's algorithm, wich takes O(nlogn) time and O(n) space. We proceeded to explain Van Leeuwen's algorithm, wich is a variation of Huffman coding and assumes that the input weights come in order. Van Leeuwen's algorithm takes O(n) time and O(n) space.

Then we explained in a detailed (and graphical) way, how does the first algorithm described by Pessoa worked. First of all, Pessoa's defined an algorithm called the "Lazy traversal algorithm" wich was a way to traverse a Huffman tree without storing the whole tree. Given this, his algorithms based on calculating the Huffman Tree in a way that this Lazy traversal algorithm worked. Pessoa's algorithm, just as Van Leeuwen, assumed that the input weights come in order.

This algorithm in particular, claimed to take O(n) time and $O(\min(H^2, n))$ space, with H the height of the Huffman Tree. After our analysis we concluded that this results where correct, at least in the paper.

We concluded that project with three "questions of interest". One of them was a satirical question about the quality of the paper, the other two was:

1. Is the use of O(H) space optimal?
2. Is it necessary for the input to be completely sorted to achieve linear time?

This second question is closely related to the work being done actually.