Abstracts

Sam Jongman & Huub Visser

7 nov. 19

Chapter 1

We have made a program that encodes a text file per line into numbers, whilst keeping special characters(!@#$%^&\*()<> etc.) intact. After the user has chosen a file, they have the option to process it and save the encoded file as a text file and save the index file. The index file is a CSV-file in which the translation (corresponding letters to the words), frequency of the words within the whole file and the amount of times a word appears in a line (abstract) is written. When it encodes a file it does it by reading the text file line by line. Each different word gets a different number assigned to it so if a word appears multiple times in a file it has the same number assigned to it. It also creates a textual histogram of the amount of times a word begins with a certain letter.

Chapter 2

Unfortunately, we did not have enough time to program a decoder and it also doesn’t calculate the TF/IDF-scores of a line.