EncodingWordsSuffixes

@author Isha Younas

@version Java 19

Description

EncodingWordsSuffixes is a Java CLI application for encoding and decoding text files using a word-to-number map. It reads a CSV file is used as a dictionary, mapping each word to a unique number. The program reads a text fil, encodes it by replacing each word with its corresponding numbers and allows decoding by reversing the process and then finally outputs the decoded text to a file specified by the user. If the program cannot encode/decode a word/number, it output "unknown ???". Designed using a command-line interface (cli) and TreeMap for an efficient and sorted order.

To Run

First compile the java files

javac -d bin src/ie/atu/sw/*.java

Then create a jar file

jar -cf dsa.jar -C bin .

and then run the program

java -cp dsa.jar ie.atu.sw.Runner

Once running, you will be prompted for the following. 1). Specify Mapping File (./encodings-10000.csv) 2). Specify Text File to Encode (./textfiles/filename.txt) 3). Specify Output File

Features

Encode text by replacing words with mapped numbers

- The program reads a text file line by line and replaces each words with its corresponding number from the csv mapping file.

Decode text by reversing the mapping

- The program reads a file with numbers and then replaces the numbers with the corresponding original word from the text files.

Input and output through the console

- All file paths are specified by the user, e.g., mapping file, file to encode and output file

Uses TreeMap for sorted and fast lookup

- A TreeMap<String, Integer> is used to store a word-to-number map, kept in a sorted order.

Handles unknown words without crashing

- If the input file contains words than are not found in the mapping file then "unknown ???" is outputted at default.