

Readme

Algorithm Description

Codes for the paper "Integrated Error Correction to Enhance Efficiency of Digital Data Storage Based on DNA Nanostructures"

File Function

clustermain.py ----- The main program of the proposed IEC algorithm. Please refer to the supplementary for the

fountainprocess.py ----- The preprocessing procedure proposed by the paper "DNA Fountain enables a robust and efficient storage architecture". Please refer to section 1.3.4 of its supplementary for details.

add_index.py ----- Add an index for each sequence.

repeat.py ----- Tally the degrees of duplication (DoD) of sequences.

error_append.py ----- (For simulation) Simulate the sequencing procedure and add errors according to the specified error rate.

Usage

Language: python3

1. (Optional) Run **repeat.py** to tally the DoD information.
2. Run **add_index.py** to add indexes.
3. Run **clustermain.py** to utilize the IEC algorithm and obtain the resulting file.
4. Input the file to the DNA fountain decoder to obtain the original file