

Documentatie Varianta Paralela

Radu Dragos Florian GR3.3A

Descriere proiect

Proiectul pe care l-am realizat se bazeaza pe metoda LZW (Lempel–Ziv–Welch) de compresie a datelor, este o metoda LOSSLESS de compresare a fisierlor text ceea ce inseamna ca nici o informatie nu este pierduta in urma acestiea.

Modul in care functioneaza acest algoritm este prin identificarea si eliminarea redundantelor din fisierele text, acesta gaseste simboluri recurente si le grupeaza in string-uri, pentru ca apoi sa le converteasca in cod, iar din moment ce codul foloseste mai putin spatiu pentru a fi stocat, fata de string-uri se realizeaza compresia.

Folosim un dictionar care verifica daca caracterul precedent (la prima iteratie este "") si caracterul prezent se afla in dictionar, daca acesta nu este regasit in dictionar atunci il stocam si caracterul prezent primeste pozitia caracterului precedent, atunci cand un caracter exista in dictionar acesta este unit cu caracterul precedent si ambele caractere devin caracter precedent.

Exemplu: "ababac" -> a -1, ab-2, ba-3, aba-4, ac-5

Exemple de rulare

Am folosit un numar de 6 procese, fiecare a primit un fisier text, fisierele text au avut 2000, 4000, 8000, 16000, 32000 si respectiv 64000 ca numar de cuvinte.

Procesul 1 – 2000 => 0.00400 seconds Procesul 5 – 32000=> 0.04800 seconds

Procesul 2 – 4000 => 0.00700 seconds Procesul 6 – 64000=> 0.08704 seconds

Procesul 3 – 8000 => 0.02500 seconds

Procesul 4 – 16000 => 0.02600 seconds

Compresia a fost realizata cu succes doar pentru ultimele doua text, celelalte au avut un numar prea mic de caractere pentru a realiza compresia in modul dorit.