

Introducció

Introduïrem a continuació una petita descripció de cada joc de proves, per a la compressió i descompressió, amb l'objectiu de veure el contingut de cada fitxer.

Fitxers txt

Els arxius de InputX.txt estan a la carpeta src/persistencia/data

Per a cada arxiu XXX.txt existeix una còpia del qual anomenada XXXCheck.txt, la qual farem servir per comprovar que el resultat obtingut de la compressió -> descompressió és el desitjat. Cal tenir en compte que si comprimim XXX.txt se'n genera XXX.jm. I si després comprimim XXX.jm se sobreescriu XXX.txt.

Recomanem la comanda "diff XXX.txt XXXCheck.txt" de Linux per comprovar el correcte funcionament del codi.

Input1.txt

Es tracta d'un fitxer de 4.4 MB en el que es repeteix un cert paràgraf un nombre determinat de vegades. Ens centrem en intentar maximitzar el grau de compressió de l'arxiu.

Input2.txt

En aquest cas estem treballant amb un fitxer de 5.5 MB, amb el que busquem veure si la velocitat de compressió dels nostres algorismes es veu greument perjudicada amb un augment de la mida o si es manté dintre d'uns marges raonables degut a que l'algorisme en qüestió segueix el grau de complexitat esperat.

Input3.txt

Tornem a tenir un fitxer de menys de 1.5 MB (1.1 MB concretament), però en aquest cas no estem repetint de manera intencionada el seu contingut. La intenció es veure com es comporten els nostres algorismes amb un text de mida raonable.

Input4.txt

Vam afegir un tercer text de mida similar al Input1/3 per comprovar si el grau de compressió es mantenia per sobre de 1 en textos de tamany superior a 1 MB.

Input5.txt

En tots els casos anteriors estàvem treballant sobre textos en català/castellà, fet que no ens acabava de convèncer per assegurar el correcte funcionament de la compressió/descompressió de textos utf8, per tant vam generar un fitxer amb codificació Koreana.

Input6.txt

A l'últim moment ens va sorgir el dubte sobre què passaria si el fitxer estava completament buit, així que vam

crear aquest sisé joc de proves.

Imatges ppm

A src/persistencia/data/imatges_grans hi ha imatges d'entre 1,4 MB i 2,3 MB.

A src/persistencia/data/imatges_petites hi ha imatges d'entre 12 kB i 232 kB.

Per a cada arxiu XXX.ppm existeix una còpia del qual anomenada XXXCHECK.ppm, la qual farem servir per comparar que el resultat obtingut de la compressió -> descompressió és el desitjat. Cal tenir en compte que si comprimim XXX.ppm se'ns genera XXX.jm. I si després comprimim XXX.jm se sobreescriu XXX.ppm.

Com que JPEG és lossy les imatges no es poden comparar amb la comanda *diff* sinó que s'ha de fer a ull.

Carpetes

src/persistencia/data/carpeta és un exemple de carpeta bastant complet pensada per ser comprimida. Conté:
- *house.ppm*, una imatge petita - *sage.ppm*, una imatge gran - *subcarpeta*, una carpeta amb:
- *curt.txt*, un fitxer de text curt - *llarg.txt*, un fitxer de text llarg - *max.cc*, el codi d'un programa en C++. Serà comprimit amb un algoritme LZ, igual que els fitxers txt.

src/persistencia/data/carpetaCHECK és una còpia de la carpeta per comprovar que el procés compressió->descompressió s'hagi fet correctament.

Junit

Els quatre fitxers que hi ha a src/persistencia/data/junit (junit.jm, junit.txt, junit_check.jm i junit_check.txt) son emprats per la comprovació de la classe LZW amb Junit. Per tal que funioni la comprovació no han de ser modificats per cap actor que no sigui el programa encarregat de fer la comprovació.