

Iseseisev töö nr 6 arendada edasi failide prefiks-pakkimise ja lahtipakkimise süsteemiks

Nõutavad (soovitavad) skeemid

Pakkimine, meetod `void pakkidaFailistFaili(String fail1, String fail2)` :

Sisestada antud nimega *fail1* (nt „NM.txt“) tekstifail sõnesse (String) *s*.

Prefiks-kodeerida see 0-1 sõnesse (String) *sKod*.

Tihendada *sKod* baidimassiivi (byte[]) *sB*:

~ iga jrk 8 sümbolit (a' '0' või '1') sõnest annavad 8 bitti (0 või 1) baiti panemiseks.

Lisada baidimassiivile *sB* veel dekodeerimiseks vajalik info ja

salvestada selliselt täiendatud massiiv *sB'* faili *fail2* (nt nimega „NM.pak“).

Printida kompressiooni protsent: $100 * (\#s - \#sB') / \#s$,

kus $\#s$ on sümbolite arv lähtesõnes *s* ja $\#sB'$ on baitide arv massiivis *sB'*.

Lahtipakkimine, meetod `String lahtiPakkidaFailist(String fail2)` :

Sisestada antud nimega *fail2* (nt „NM.pak“) baidifail baidimassiivi (byte[]) *sB'*.

Kasutades selles olevat dekodeerimiseks vajalikku infot,

dekodeerida selles olev prefikskoodide osa sõneks *sDekod*; tagastada see sõne.

Ülesande võib kitsendada juhule, kus eeldatakse, et lähte-tekstifailis *fail1* ei ole sümboleid suurema koodiga kui 255; tähtedest – ladina tähestik + eesti täpitähed. Tüüpiliselt on sellisteks failideks nt Java-programmide lähtekoodid. Üks piiratud sümbolivaruga näitefailide komplekt leidub arvutipraktikumide materjalides.

Kõrvaolev tabel iseloomustab ühte olemasolevat, kitsendatud sümbolivaruga lähte-tekstifailide prefiks-pakkimise süsteemi.

Fail	Maht	Pakitu maht	Kompressioon
-----	-----	-----	-----
LoremIpsum.txt	4000	2226	44%
Jutte_I_Ojam.txt	8331	4874	41%
UlmelineRiismaa_P_Pöldver.txt	9990	5915	40%
RiiaAknad_K_Aareleid.txt	12504	7241	42%
LoremIpsum+.txt	29041	15674	46%