

## KOMPRESIJA I ZAŠTITA PODATAKA

### - Projektni zadatak 1 -

Potrebno je izračunati bajt-entropiju binarnog fajla i implementirati osnovne algoritme za kompresiju.

1. Izračunati bajt-entropiju datog binarnog fajla. Označimo sa  $N_i$  broj pojavljivanja bajta  $i = 0, 1, \dots, 255$  u datom binarnom fajlu, kao i  $p_i = N_i/N$ , gde je  $N$  ukupna dužina fajla u bajtovima. Bajt-entropija je definisana izrazom

$$H(p) = -p_0 \log_2 p_0 - p_1 \log_2 p_1 - \dots - p_{255} \log_2 p_{255}.$$

Pritom podrazumevamo da je  $0 \log_2 0 = 0$ .

2. Konstruisati Shannon-Fano i Huffmanov kod na osnovu vrednosti  $p_0, p_1, \dots, p_{255}$  i primeniti ih na kodiranje (odnosno kompresovanje) datog binarnog fajla. U kodiranom fajlu, potrebno je najpre zapamtiti sam kod, a zatim i kodirane podatke iz ulaznog fajla.
3. Implementirati algoritme LZ77 i LZW i primeniti ih na kompresiju datog binarnog fajla. Pretpostaviti da je skup simbola ulaznog alfabeta  $A = \{0, 1, \dots, 255\}$ .

Za maksimalni broj poena, potrebno je osmisлити strukturu kodiranog fajla, tako da se postiže (asimptotski) optimalna veličina fajla. Imajte u vidu da se kodirani podaci predstavljaju nizom bitova (naročito kod Shannon-Fano i Huffmanovog koda), pa je za optimalno skladištenje, potrebno memorisati 8 bita po bajtu.

Potrebno je implementirati i proces dekodiranja za svaki navedeni kod. Jedan od načina da testirate funkcionalnost koda je da fajl kodirate, zatim dekodirate i onda poredite sa originalnim fajlom (za ovu priliku mogu poslužiti command-line alati `fc` i `diff`, na operativnim sistemima Windows odnosno Linux (i MacOS).

Odabрати binarni ili tekstualni fajl veličine oko 1-10MB, primeniti algoritme i odrediti stepen kompresije svakog od metoda. Napisati kratak izveštaj u obliku `txt` fajla, zajedno sa kratkim opisom kako se pokreće implementacija.

Implementacije obaviti u nekom od jezika: C++, Java, Python, C#. Zabranjeno je korišćenje nestandardnih biblioteka.

May the Force be with you!!!

Predmetni nastavnik  
dr Marko Petković, red. prof.