

# Teorija informacija - laboratorijska vježba, zadatak 3.

Ovo je C# implementacija Shannon-Fano koder za kolegij Teorija informacija 5. semestra FERovog preddiplomskog studija. Slobodno smijete koristiti bilo koji dio izvornog koda sve dok mi to dojavite. Kod je isporučen kakav je i autor osobno ne odgovara za bilo koju štetu nastalu korištenjem dijela ovog koda ili koda kao cijeline.

## Ulaz

Postoje dva tipa ulaza: uzorak i opis. Ulazna točka nalazi se u EntryPoint.cs.

### Ulaz: uzorak

Korisnik upisuje niz znakova koji predstavljaju statistiku modela ulaznih znakova. Na primjer, ako korisnik upiše "Hakuna matata" kao ulaz, to će biti jednako sljedećim vjerojatnostima znakova:

```
<a> 5/13
<t> 2/13
< > 1/13
<H> 1/13
<k> 1/13
<m> 1/13
<n> 1/13
<u> 1/13
```

### Ulaz: opis

Korisnik upisuje opis vjerojatnosti svakog znaka u formatu sličnom formatu iznad. Slijede pravila:

- Svaki stohastički element je odvojen točka-zarezom (;).
- Svaki stohastički element predstavljen je točno jednim znakom za vrijednost i brojem koji opisuje vjerojatnost pojave navedenog stohastičkog elementa.
- Svaka vrijednost i vjerojatnost odvojeni su točno jednim razmakom ().
- Svaka vrijednost piše se kao <vrijednost>. Drugim riječima, vrijednost je sadržana između < i >.
- Svaka vjerojatnost piše se ili kao realan broj (format 0.000... ili 0,000...) ili kao razlomak N+brojeva (format 0/0).

Ispod se nalazi primjer ovakvog ulaza

```
<a> 0.5;
<b> 0,25;
<c> 1/4
```

Također, s obzirom da je provjera unosa obavezna, uvedena je i normalizacija ulaza - ako program uoči da suma vjerojatnosti slučajne varijable nije 1, korisnik će dobiti upit želi li normalizirati slučajnu varijablu. Ako odgovori potvrdno, program će svaku vrijednost podijeliti sumom vjerojatnosti, tako da će nova suma vjerojatnosti uvijek biti što jednakija jedan (neki put to neće biti moguće zbog nesavršenosti zapisa decimalnih brojeva u računarstvu, iako se koriste 128-bitne vrijednosti. Ako korisnik odgovori negativno, program će ispisati poruku da suma vjerojatnosti nije jednaka 1 te će pitati korisnika za novi unos.

## Konačne riječi vezane uz unos podataka

S obzirom da ulazna točka čita i ulaz u konzolu i sadržaj datoteka koje se prilože na standardni ulaz, prihvaća i EOF i "" na nekoj liniji kao kraj unosa. Dok ovo znači da će se prilikom učitavanja datoteke rezultati ispisati odmah, učitavanje podataka preko komandne linije će zahtijevati da se za kraj upiše prazna linija nakon unosa podataka. Ovo obično znači da će se na kraju samo dva puta pritisnuti return.

Jednostavnije, ako korisnik želi upisati abcde kroz konzolu, trebat će upisati:

```
abcde
```

## Ispis

Pravila ispisa su:

- Prvo se ispisuje unost podataka, ispisan sortirano s padajućim vjerojatnostima.
- Nakon originalnog ispisa, koder izrađuje svoj nacrt kodova za svaki stohastički element, prikazujući parove vrijednosti i pripadnog koda.
- Naposljetku, ispisana je srednja kodna duljina, kao i efikasnost Shannon-Fano koda.

Primjer ispisa za abcde je:

```
The given description is as follows:
```

```
<a>    0,200
<b>    0,200
<c>    0,200
<d>    0,200
<e>    0,200
```

```
For the given description, the Shannon Fano code is:
```

```
-----
<a>    ->    00
<b>    ->    01
```

```
<c>    ->    10
<d>    ->   110
<e>    ->   111
```

Mean code word length = 2,400  
Code efficiency: 96,747%

-----

Ispis za nešto kompliciraniji niz znakova, koji je

Peter Piper picked a a peck of pickled peppers, a peck of picked peppers Peter Piper Picked. If Peter Piper picked a peck of pickled peppers, where's the peck of pickled peppers Peter Piper picked?

bi nakon normalizacije bio:

The given description is as follows:

```
<e>    0,178
< >    0,173
<p>    0,137
<r>    0,066
<c>    0,061
<i>    0,061
<k>    0,061
<P>    0,046
<d>    0,041
<f>    0,025
<s>    0,025
<t>    0,025
<a>    0,020
<o>    0,020
<l>    0,015
<, >  0,010
<h>    0,010
<'>    0,005
<.>    0,005
<?>    0,005
<I>    0,005
<w>    0,005
```

For the given description, the Shannon Fano code is:

-----

```
<e>    ->    00
< >    ->    010
<p>    ->    011
<r>    ->   1000
<c>    ->   1001
<i>    ->   1010
<k>    ->   1011
<P>    ->   11000
<d>    ->   11001
```

```
<f>    ->    11010
<s>    ->    11011
<t>    ->    11100
<a>    ->    111010
<o>    ->    111011
<l>    ->    111100
<,>    ->    111101
<h>    ->    1111100
<'>    ->    1111101
<.>    ->    11111100
<?>    ->    11111101
<I>    ->    11111110
<w>    ->    11111111
```

Mean code word length = 3,756

Code efficiency: 98,836%

-----

## Izvršavanje

Za izvršavanje je potreban najnoviji .NET Framework, a izvorni kod je u C# 7.1.