

Programarea calculatoarelor și limbaje de programare I

Tema 2

Termen de predare: săptămâna 8-12 ianuarie 2024

Punctaj: ETTI 10%, CALC 15% din nota finală

Entropia informațională a fost introdusă în anul 1948 de Claude Shannon și este o metrică prin care se calculează informația medie dintr-un mesaj. Această noțiune se bazează pe ideea că un eveniment mai frecvent, deci cu probabilitate de apariție mai mare, înglobează mai puțină informație decât un eveniment mai rar, deci cu o probabilitate mai mică. Entropia se notează de obicei cu litera H și se calculează astfel:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

unde x_i este unul dintre evenimentele x_1, x_2, \dots, x_n care sunt realizări ale variabilei aleatoare X , iar $p(x_i)$ este probabilitatea de apariție a evenimentului x_i . Dacă folosim logaritmul în baza 2 în formula entropiei, atunci unitatea sa de măsură a entropiei se numește *bit*. Dacă folosim logaritmul natural, unitatea de măsură a entropiei se numește *nat*.

Atunci când ne stabilim o parolă nouă, dorim ca aceasta să fie cât mai puternică și mai puțin predictibilă. Entropia poate fi folosită pentru a calcula varietatea de caractere dintr-un text. Cu cât entropia unei parole este mai mare, cu atât ea este mai puternică și mai greu de produs. De exemplu, entropia textului 7364472 este 2,236, iar entropia textului 7364472599 este 2,722.

Entropia unui text se calculează astfel :

- Se calculează probabilitatea de apariție a fiecărui caracter în textul examinat prin împărțirea numărului de apariții ale fiecărui caracter la lungimea textului. Pentru textul 7364472, probabilitatea de apariție a cifrelor 2, 3 și 6 este $1/7 \approx 0,142$, iar a cifrelor 4 și 7 este $2/7 \approx 0,285$.
- Se calculează entropia textului folosind formula entropiei. Entropia textului de mai sus este $-3 \cdot \frac{1}{7} \cdot \log_2 \frac{1}{7} - 2 \cdot \frac{2}{7} \cdot \log_2 \frac{2}{7} = 2.236$.

Puteti să calculați entropia unui text folosind calculatorul online disponibil la adresa <https://onlinetexttools.com/calculate-text-entropy>. Generați diverse parole cu <https://www.avast.com/random-password-generator>, copiați-le în calculatorul entropiei și verificați cum se modifică entropia odată cu complexitatea parolei.

Scriveți un program C++ care calculează entropia unei parole citite de la tastatură sau dintr-un fișier și care este formata doar din cifre de la 0 la 9.

Sugestie de implementare: Cifrele sunt citite una câte una într-o variabilă de tip `char`, iar numărul de apariții ale fiecărei cifre este păstrat într-un tablou cu 10 elemente. La apariția cifrei i se incrementează elementul i al tabloului.

Predarea temei se va face la orele de laborator ale fiecărei subgrupe.