**Curs**: Analiza statistică a datelor

**Tema**: Probabilitatea condițională și entropia

**Sarcina**: De calculat entropia medie per simbol pentru textul dat în varianta respectivă

**Varianta 11:**

1. Deodată, furnicuţa Observă că zboară-n sus

Litera a, și litera a precedată de litera ț.

**Rezolvarea**: in textul dat sunt 42 caractere (cu spații) și 3 litere **a**.

Probabilitatea literei **a** în textul dat este P(a) = 3/42 = 0.0714.

Entropia, respectiv H(a) = -P(a) log2(P(a)) = -0.0714 log2(3/42) = -0.0714 (-3.807) = 0.2719

Astfel, entropia literei **a** este 0,2719

Entropia condiționată este calculată în baza probabilității condiționate.

H(a/ț) = -P(a/ț) log2(P(a/ț))

Probabilitatea condiționată se calculează cu formula

P(a/ț) = P(ț,a)/P(ț)

In textul dat este 1 caz când litera **a** este precedată de litera **ț**. P(ț,a) = 1/42 = 0.0238

Probabilitatea literei ț în textul dat este P(ț) = 1/42 = 0.0238 deoarece apare în textul dat doar de 1 ori.

Respectiv, P(a/ț) = P(ț,a)/P(ț) = 1/42 / 1/42 = 1

Entropia, respectiv H(a/ț) = -P(a/ț) log2(P(a/ț)) = -1 log2(1) = 0

În cazul dat probabilitatea este 1, ce înseamnă eveniment sigur și nu este nici o incertitudine. Entropia, ca măsura incertitudinii, în cazul dat este egală cu 0.