**Curs**: Analiza statistică a datelor

**Tema**: Probabilitatea condițională și entropia

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

**Varianta 13:**

1. Lumea largă – hopa-hop – Vrea s-o vadă

Litera o, și litera o precedată de litera h.

**Rezolvarea**: in textul dat sunt 38 caractere (cu spații) și 3 litere **o**.

Probabilitatea literei **o** în textul dat este P(a) = 3/38 = 0.0789.

Entropia, respectiv H(a) = -P(a) log2(P(a)) = -0.0789 log2(3/38) = -0.0789 (-3.664) = 0.289

Astfel, entropia literei **o** este 0,289

Entropia condiționată este calculată în baza probabilității condiționate.

H(o/h) = -P(o/h) log2(P(o/h))

Probabilitatea condiționată se calculează cu formula

P(o/h) = P(h,o)/P(h)

In textul dat este 2 cazuri când litera **o** este precedată de litera **h**. P(h,o) = 2/38 = 0.052

Probabilitatea literei h în textul dat este P(h) = 2/38 = 0.052 deoarece apare în textul dat doar de 2 ori.

Respectiv, P(o/h) = P(h,o)/P(h) = 2/38 / 2/38 = 1

Entropia, respectiv H(o/h) = -P(o/h) log2(P(o/h)) = -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.