

# Segurança de Sistemas – Índice de Coincidência

Prof. Avelino F. Zorzo - Escola Politécnica PUCRS

Através do índice de coincidência é possível identificar a língua na qual um texto foi escrito. Esta atividade será utilizada no primeiro trabalho.

O objetivo desta atividade é desenvolver um programa (em qualquer linguagem de programação) que:

- calcule a frequência  $f_i$  de cada caractere (a..z) em um determinado texto ( $f_0$  é a frequência do caractere **a**, ou seja, quantos caracteres **as** existem no texto e assim por diante –  $n$  é o número de caracteres do texto)
- calcule o índice de coincidência de um texto segundo a fórmula abaixo (valor para um caractere):

$$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n - 1)} \approx \sum_{i=0}^{25} p_i^2$$

Após o cálculo do índice de coincidência, verificar os textos cifrados T1.txt ... T7.txt disponibilizados no Moodle e dizer

- quais textos cifrados são iguais;
- qual a provável língua do texto claro.

As tabelas de frequência das letras em português/inglês/... podem ser encontradas em:  
[https://en.wikipedia.org/wiki/Letter\\_frequency](https://en.wikipedia.org/wiki/Letter_frequency).

Tabela 1 – Índices de coincidência

Língua	Índice de coincidência <sup>1</sup>
Alemão	2,05 (0,078)
Espanhol	1,94 (0,074)
Francês	2,02 (0,077)
Inglês	1,73 (0,066)
Italiano	1,94 (0,074)
Português	1,94 (0,074)
Aleatório	1,00 (0,0385)

**Pergunta: Existe algum dos textos cifrados que pode ser o texto claro TextoClaro.txt cifrado?**

Responder questionário no Moodle.

BOM TRABALHO.

---

<sup>1</sup> Valores aproximados.