

Algoritmo para o calculo de entropia dos dados de um arquivo de texto

```
function entropia_de_textos(arquivo)

    % Le o arquivo e adiciona seu texto a uma string
    fid = fopen(arquivo, 'r');
    texto = fscanf(fid, '%c');
    fclose(fid);

    % Converte a string em um vetor
    vetor = char(texto);

    % Cria um vetor e coloca todos os caracteres nele sem repeticao
    caracteres = unique(vetor);

    % Cria um vetor para atribuir a quantidade de ocorrencias de cada
    % caractere
    ocorrencias = zeros(size(caracteres));
    probabilidade = zeros(size(caracteres));
    % Conta o numero de ocorrencias de cada caractere
    % Lembrete: o primeiro valor a ser contado o caractere nulo, a cada
    % quebra de linha existe um caractere nulo
    for i = 1:length(caracteres)
        ocorrencias(i) = sum(vetor == caracteres(i));
        probabilidade(i) = ocorrencias(i) / length(vetor);
    end
    %Exibe os resultados
    disp('Caractere | Ocorrencias | Probabilidade');
    disp('-----');
    for i = 1:length(caracteres)
        fprintf('%9s | %d | %.2f\n', caracteres(i), ocorrencias(i), probabilidade(i));
    end
    fprintf("Entropia: %.2f\n", entropia(probabilidade));

end
```

Dado um arquivo de texto, o algoritmo mostrar todos os caracteres e a probabilidade de cada um ocorrer

Exemplo:

Um arquivo com uma linha de texto contendo "123456782444666668888888", o algoritmo mostrar na tela quantas vezes cada caractere ocorreu, e qual a probabilidade dele ocorrer, em seguida mostrar a entropia de todo o conjunto

Neste caso o algoritmo mostrar:

Caractere	Ocorrencias	Probabilidade
1	1	0.04
2	2	0.08

3		1		0.04
4		4		0.17
5		1		0.04
6		6		0.25
7		1		0.04
8		8		0.33

Entropia: 2.52