



Pontifícia Universidade Católica de Minas Gerais

Instituto de Ciências Exatas e Informática

Algoritmos e Estruturas de Dados III | Prof. Hayala Curto

Lista de Exercícios 2

1 ponto

A lista deve ser feita **individualmente** ou em grupos de no **máximo 2 alunos**.

Penalidade por atraso: a cada dia corrido de atraso, a nota será penalizada em 0,5 ponto.

Penalidade por cópia: trabalhos iguais não são aceitos (nota 0).

Parte teórica (0,5 ponto)

Lista Invertida

1 - Em qual contexto é recomendado usar lista invertida?

2 - Quais as vantagens do uso de índices invertidos sobre as outras formas de índices?

3 - O que são os termos não significativos (*stop words*) que devem ser retirados do conjunto de termos indexados em índices invertidos?

4 - Descreva um exemplo de um contexto do uso de arquivos em que a lista invertida seria importante. Mostre a entidade do seu exemplo, os campos e a aplicação da lista invertida.

Compressão de dados

1 - Uma compressão sem perdas é caracterizada por?

2 - Calcule a entropia (o tamanho médio em bits dos símbolos) da seguinte mensagem:

A ARANHA ARRANHA A RÃ

3 - Quantos bits são necessários para se representar a mensagem abaixo usando a codificação de Huffman, Shannon Fano, LZ77, LZ78 e LZW?

A ARANHA ARRANHA A RÃ

4 - Quantos bits são necessários para se representar a mensagem abaixo usando a codificação de Huffman, Shannon Fano, LZ77, LZ78 e LZW?

BOTE A BOTA NO BOTE E TIRE O POTE DO BOTE



Pontifícia Universidade Católica de Minas Gerais

Instituto de Ciências Exatas e Informática

Algoritmos e Estruturas de Dados III | Prof. Hayala Curto

5 - Considere que uma determinada implementação do LZW usa o seguinte dicionário inicial, em que o símbolo \square (posição 27) representa o espaço em branco e que possui 1024 posições.

1	A	9	I	17	Q	25	Y
2	B	10	J	18	R	26	Z
3	C	11	K	19	S	27	\square
4	D	12	L	20	T		
5	E	13	M	21	U		
6	F	14	N	22	V		
7	G	15	O	23	W		
8	H	16	P	24	X		

Usando esse dicionário inicial, decodifique a seguinte mensagem:

13, 1, 28, 5, 27, 28, 14, 4, 15, 21, 32, 21, 4, 1, 18

6 - Considere que uma determinada implementação do LZW usa o seguinte dicionário inicial, em que o símbolo \square (posição 27) representa o espaço em branco e que possui 1024 posições.

1	A	9	I	17	Q	25	Y
2	B	10	J	18	R	26	Z
3	C	11	K	19	S	27	\square
4	D	12	L	20	T		
5	E	13	M	21	U		
6	F	14	N	22	V		
7	G	15	O	23	W		
8	H	16	P	24	X		

Usando esse dicionário inicial, codifique a seguinte mensagem:

MARA AGARRA E AMARRA A ARARA

Represente os valores separando-os por um único espaço em branco (ex.: 13 10 20 ...)

7 - Quantos bits são necessários para se representar a mensagem abaixo usando a codificação de Huffman, Shannon Fano, LZ77, LZ78 e LZW?

O_PEITO_DO_PÉ_DO_PEDRO_É_PRETO.



Pontifícia Universidade Católica de Minas Gerais

Instituto de Ciências Exatas e Informática

Algoritmos e Estruturas de Dados III | Prof. Hayala Curto

8 - Quantos bits são necessários para se representar a mensagem abaixo usando a codificação de Huffman, Shannon Fano, LZ77, LZ78 e LZW?

A_BANDA_DOS_URUBUS_BATUCA_MUITO

Casamento de padrões

1 - Considere o primeiro teste de busca do padrão CANOA no texto abaixo, usando o algoritmo Boyer-Moore:

IA NA CANOA, RIO ABAIXO

Qual será o deslocamento por caráter ruim nesse teste específico?

2 - Considere o primeiro teste de busca do padrão CANOA no texto abaixo, usando o algoritmo Boyer-Moore:

IA NA CANOA, RIO ABAIXO

Qual será o deslocamento por sufixo bom nesse teste específico?

3 - Qual é a principal vantagem do algoritmo de busca KMP sobre o algoritmo da força bruta?

4 - Qual é a representação correta do vetor de transições de falhas do padrão ABCABCAC no algoritmo KMP?

5 - Se criarmos um diagrama de estados para a busca dos termos abaixo por Aho-Corasick, esse diagrama de estados conterá quantos estados?

Termos de busca:

FACA, FOICE, CABO, CORTE

6 - Qual é a complexidade (de tempo) de um casamento de padrões por Aho-Corasick, considerando a busca de m padrões em um texto de n caracteres?

7 - Qual é a distância de edição entre os padrões CULTURA e SUTURAS?



Pontifícia Universidade Católica de Minas Gerais

Instituto de Ciências Exatas e Informática

Algoritmos e Estruturas de Dados III | Prof. Hayala Curto

8 - Qual será o valor da célula destacada em vermelho, no quadro de casamento aproximado de padrões abaixo, usando o cálculo de distância de Levenshtein?

	S	A	C	I
	0	1	2	
C	1	1	2	
A	2	2		
P				
I				
M				

9 - Usando o algoritmo de Boyer-Moore, faça a busca de TODAS as ocorrências do padrão IARA no texto abaixo (não pare na primeira ocorrência, você precisa testar toda o texto).

A IARA AMARRA A ARARA DE ARARAQUARA

Represente todos os testes e, a cada um deles, indique o deslocamento por caráter ruim (DCR) e o deslocamento por sufixo bom (DSB) calculados e circule aquele que foi usado. Lembre-se que, quando o padrão for encontrado, o deslocamento é de apenas 1 posição.

Parte Implementação (0,5 ponto)

1 - Implemente os seguintes algoritmos:

- Compactação:
 - Huffman
 - LZW
- Casamento de padrões
 - Boyer Moore