Relatório Trabalhos Práticos AEDS3

Felipe Carvalho, João Vitor Baldo

Ciência da Computação - PUC Minas

Resumo

Neste relatório fala sobre como foram feitos os trabalhos práticos para a matéria Algoritmos e Estruturas de Dados 3. Para o trabalho prático 1 era necessário implementar CRUD Sequencial e Ordenação Externa. Para o trabalho prático 2 era necessário implementar indexação utilizando Árvore B, Hashing Estendido e Lista Invertida. Para o trabalho prático 3 era necessário fazer a implementação dos algoritmos Huffman, LZW e dois algoritmos de casamento de padrões Para o último trabalho prático era necessário implementar dois algoritmos de criptografía.

Trabalho Prático 1 - Criação da base de dados, Manipulação de Arquivo Sequencial e Ordenação Externa

No primeiro trabalho prático da disciplina, desenvolvemos uma aplicação para manipular e organizar informações dos personagens dos filmes da saga Harry Potter.

O trabalho começou com a implementação de uma classe para realizar a leitura de um arquivo CSV contendo os dados dos personagens, como identificador, nome, data de nascimento, e outros atributos. Essas informações são armazenadas em um arquivo binário com o seguinte formato:

- Um número inteiro no início do arquivo para registrar o último ID inserido.
- Uma lápide de 1 byte que indica se o registro é válido (1 para válido e 0 para excluído).
- Um inteiro representando o tamanho (em bytes) do registro do personagem.
- Os dados do personagem são armazenados em formato estruturado, incluindo inteiros, strings e datas (representadas como valores long).

Foi implementado também operações básicas de CRUD (Create, Read, Update e Delete), permitindo a manipulação completa dos registros de personagens.

Por fim, foi desenvolvido um algoritmo de ordenação externa para reorganizar os registros no arquivo binário. Este algoritmo foi projetado para lidar com grandes volumes de dados, que excedem a capacidade de memória principal, utilizando estratégias de divisão, ordenação em blocos menores e intercalamento.

Trabalho Prático 2: Manipulação de Arquivo Indexado com Árvore B, Hash e Lista Invertida

No segundo trabalho prático, foi implementado a inclusão de estruturas de dados avançadas para otimizar as operações de busca e manipulação no arquivo de dados. Foram implementadas três abordagens principais: Árvores B, Hashing Estendido e Listas Invertidas.

-Árvore B de Ordem 8 com Arquivo de Índice

Foi implementada uma Árvore B de ordem 8, acompanhada de um arquivo de índice para armazenar os pares "ID do personagem, posição no arquivo de dados".

A Árvore B foi projetada para realizar buscas eficientes no arquivo binário de dados, utilizando o índice para localizar rapidamente a posição de um personagem específico. Sempre que havia alterações no arquivo de dados, como inserções, exclusões ou atualizações, o arquivo de índice era automaticamente atualizado para refletir essas mudanças. Garantindo um desempenho otimizado mesmo em grandes volumes de registros.

-Hashing Estendido com Arquivo de Índice

Também foi implementado um Hashing Estendido utilizando um arquivo de índice que mapeava o ID do personagem e posição no arquivo de dados.

A função hash utilizada foi $h(k) = k \mod 2^p$, onde p representa a profundidade do diretório. Cada bucket podia armazenar até X registros, sendo X equivalente a 2% do tamanho inicial da base (404 registros). Sempre que um bucket atingia sua capacidade máxima, acontecia um split no diretório, adaptando a estrutura para comportar novos registros. O índice era atualizado em tempo real sempre que o arquivo de dados sofria alterações, garantindo a

consistência entre os dois arquivos. O objetivo do Hashing Estendido era fornecer uma alternativa eficiente para buscas rápidas no arquivo de dados.

-Lista Invertida

E por último, foi implementado um sistema de Listas Invertidas para permitir buscas mais flexíveis no arquivo binário de dados.

As listas invertidas armazenavam informações de índices secundários, associando atributos não primários, como ano de nascimento ou casa do personagem, às posições dos registros no arquivo de dados. Sempre que um registro era inserido, alterado ou deletado no arquivo de dados, as listas invertidas eram atualizadas para refletir essas mudanças. O sistema foi projetado para permitir buscas que utilizassem múltiplas listas invertidas simultaneamente, possibilitando pesquisas complexas que combinavam diferentes critérios, como, por exemplo, personagens de uma determinada casa e com uma data de nascimento específica.

Trabalho Prático 3: Compactação e Casamento de Padrões

No terceiro trabalho prático, foi implementado algoritmos de compactação de dados e casamento de padrões. O objetivo foi tornar o sistema mais eficiente no armazenamento e processamento de informações, utilizando algoritmos consagrados para cada tarefa.

-Compactação do Arquivo Binário

Foi adicionada ao programa a capacidade de compactar o arquivo binário de dados, permitindo ao usuário escolher quando realizar esse processo. Para isso, foram utilizados dois algoritmos de compactação:

 Huffman: Baseado em árvores binárias, esse algoritmo atribui códigos menores a caracteres mais frequentes, maximizando a eficiência da compactação. LZW: Utiliza um dicionário dinâmico para identificar padrões repetitivos no arquivo e reduzir seu tamanho de forma eficiente.

Após a compactação, o programa oferece ao usuário a opção de descompactar o arquivo. Quando essa operação era realizada, o arquivo descompactado substitui o arquivo arquivo de dados e passa a ser utilizado como o novo arquivo de dados pelo programa, garantindo que as operações subsequentes fossem realizadas sobre os dados descompactados.

O sistema também apresentava ao usuário métricas importantes, como:

- O tempo de execução dos algoritmos Huffman e LZW durante a compactação.
- A porcentagem de perda após a descompactação comparando o tamanho com o antigo arquivo de dados.

-Casamento de Padrões

Foi implementada ao Read a funcionalidade de busca de padrões nos registros dos personagens, utilizando os algoritmos KMP e Rabin-Karp.

O usuário fornece o ID do personagem e o padrão que deseja encontrar no registro do personagem e o programa realiza a busca utilizando os dois algoritmos, no final o programa mostra para o usuário se o padrão desejado foi encontrado ou não nos dados do personagem.

Trabalho Prático 4: Criptografia

O quarto trabalho prático teve objetivo a implementação de técnicas de criptografía, adicionando ao programa funcionalidades para proteger os dados armazenados e processados. Foram aplicados dois algoritmos amplamente utilizados no campo da segurança da informação: RSA e Cifra Sequencial.

O programa foi configurado para executar as operações de criptografia e descriptografia e quando o usuário digitar "Encriptar" no console. Quando ativado, ele realiza a criptografia e a descriptografia do arquivo de dados gerando no total 4 arquivos, sendo 2 criptografados pelos algoritmos RSA e cifra por substituição e 2 arquivos descriptografados.

-RSA

Foi implementado o algoritmo RSA, que utiliza a criptografía assimétrica para proteger os dados.

O programa utiliza uma chave pública para fazer a criptografía dos dados e uma chave privada que necessita ser mantida em sigilo para a fazer descriptografía dos dados.

Ao digitar "Encriptar" o programa realiza a criptografía dos dados e em seguida realiza a descriptografía dos dados.

-Cifra por Substituição

Foi implementada também a Cifra por Substituição, uma técnica de criptografía simétrica simples que substitui cada caractere por outro de acordo com uma regra definida.

Para a criptografia o programa aplica uma chave de substituição predefinida, onde o byte do arquivo é substituído pelo valor na tabela.

Para a descriptografia o programa faz o processo inverso gerando o arquivo igual o arquivo de dados.

Testes -Arquivo CSV de origem

```
| disease_literate_nees:|suse_ascesty;species;patronus;logoartsStaff;hogoartsStudent;sctorName,allve;alternate_ascors;dateOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;varOfBirth;va
```

Todos os algoritmos feitos têm este csv como base e realizam operações nos registros levando em conta os tipos de seus atributos.

-Arquivo Binário

O arquivo binário gerado que tem a seguinte formatação:

- Um número inteiro no início do arquivo para registrar o último ID inserido.
- Uma lápide de 1 byte que indica se o registro é válido (1 para válido e 0 para excluído).
- Um inteiro representando o tamanho (em bytes) do registro do personagem.
- Os dados do personagem são armazenados em formato estruturado, incluindo inteiros,

strings e datas (representadas como valores long).

```
£33
         00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
                                                    Decoded Text
        00 00 01 93 00 00 00 00 E7 00 00 00 00 00 24 39
        65 33 66 37 63 65 34 2D 62 39 61 37 2D 34 32 34
                                                     e 3 f 7 c e 4 - b 9 a 7
        34 2D 62 37 30 39 2D 64 61 65 35 63 31 66 31 64
                                                            0 9
                                                                  d a
00000030 34 61 38 00 0C 48 61 72 72 79 20 50 6F
00000040 72 00 3D 5B 54 68 65 20 42 6F 79 20 57 68 6F 20
00000050 4C 69 76 65 64 2C 20 54 68 65 20 43 68 6F 73 65
000000000 6E 20 4F 6E 65 2C 20 55 6E 64 65 73 69 72 61 62
                                                                  Undes
00000070 6C 65 20 4E 6F 2E 20 31 2C 20 50 6F 74 74 79 5D
000000080 00 0A 47 72 79 66 66 69 6E 64 6F 72 00 0A 68 61
                                                         Gryffindor
                                                     1 f - b 1 o o d . .
000000090 6C 66 2D 62 6C 6F 6F 64 00 05 68 75 6D 61 6E 00
                                                                        human
000000A0 04 73 74 61 67 00 05 66 61 6C 73 65 00 04 74 72
                                                       stag..
                                                                  false
00000080 75 65 00 10 44 61 6E 69 65 6C 20 52 61 64 63 6C
                                                            Daniel
                                                                          Radcl
00000000 69 66 66 65 00 04 74 72 75 65 00 00 00 4D BB A1
                                                     iffe...
                                                                true
000000D0 6B 80 00 05 67 72 65 65 6E 00 04 6D 61 6C 65 00
                                                            green.
                                                                          male
000000E0 05 62 6C 61 63 6B 00 04 74 72 75 65 00 00 07 BC
                                                       black.
        00 00 00 00 DC 00 00 00 01 00 24 34 63 37 65 36
00000100 38 31 39 2D 61 39 31 61 2D 34 35 62 32 2D 61 34
00000110 35 34 2D 66 39 33 31 65 34 61 37 63 63 65 33 00
                                                          f 9 3 1 e 4 a 7
00000120 10 48 65 72 6D 69 6F 6E 65 20 47 72 61 6E 67 65
00000130 72 00 30 5B 48 65 72 6D 79 2C 20 4B 6E 6F 77 2D
                                                         0[Hermy,
00000140 69 74 2D 61 6C 6C 2C 20 4D 69 73 73 20 47 72 61
                                                     it-all,
                                                                    Miss
00000150 6E 74 2C 20 48 65 72 6D 2D 6F 77 6E 2D 6E 69 6E
                                                            Herm-own
00000160 6E 79 5D 00 0A 47 72 79 66 66 69 6E 64 6F 72 00
                                                     ny]..Gryffindor
muggleborn.
00000180 61 6E 00 05 6F 74 74 65 72 00 05 66 61 6C 73 65
                                                            otter..
                                                                          false
00000190 00 04 74 72 75 65 00 0B 45 6D 6D 61 20 57 61 74
                                                         true. . Emma
        73 6F 6E 00 04 74 72 75 65 00 00 00 47 60 47 DB
                                                     son.
                                                              true.
000001B0 80 00 05 62 72 6F 77 6E 00 06 66 65 6D 61 6C 65
                                                           brown.
000001C0 00 05 62 72 6F 77 6E 00 04 74 72 75 65 00 00 07
                                                         brown.
                                                                      tr
000001E0 66 39 61 35 2D 62 38 37 62 2D 34 38 62 66 2D 62
                                                        a 5 - b 8 7 b
        30 30 64 2D 39 35 62 30 39 33 65 61 36 33 39 30
        00 0B 52 6F 6E 20 57 65 61 73 6C 65 79 00 54 5B
                                                         Ron
```

-LZW

Após fazer a compactação utilizando o algoritmo LZW, o programa gera uma arquivo com os dados compactados

TP3 > ≣ (chara	icte	rsLz	ZWC	com	pres	sao	1.bii	n								
63	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded Text
00000000	00	01	09	EA	00	00	00	00	00	00	00	00	00	00	00	01	
00000010	00	00	00	93	00	00	01	00	00	00	01	00	00	00	00	E7	
00000020	00	00	01	04	00	00	01	00	00	00	00	24	00	00	00	39	\$ 9
00000030	00	00	00	65	00	00	00		00	00	00	66	00	00	00		
00000040	00	00	00	63	00	00	00	65	00	00	00	34	00	00	00	2D	
00000050	00	00	00	62	00	00	00	39	00	00	00	61	00	00	00		b 9 a 7
00000060	00	00	00	2D	00	00	00	34	00	00	00	32	00	00	00	34	4 2 4
00000070	00	00	01	11	00	00	00	62	00	00	00		00	00	00	30	7 0
00000080	00	00	00	39	00	00	00	2D	00	00	00	64	00	00	00	61	
00000090	00	00	00	65	00	00	00	35	00	00	00	63	00	00	00	31	
000000A0	00	00	00	66	00	00	00	31	00	00	00	64	00	00	00	34	
000000В0	00	00	00	61	00	00	00	38	00	00	00	00	00	00	00	0C	
000000C0	00	00	00	48	00	00	00	61	00	00	00	72	00	00	00	72	H a r r
00000000	00	00	00	79	00	00	00	20	00	00	00	50	00	00	00	6F	y P o
000000E0	00	00	00	74	00	00	00	74	00	00	99	65	00	00	00	72	
000000F0	00	00	00	00	00	00	00	3D	00	00	00	5B	00	00	00	54	
00000100	00	00	00	68	00	00	00	65	00	00	00	20	00	00	00	42	h e B
00000110	00	00	00	6F	99	99	01		99	00	00	57	99	99	00	68	o 3 W h
00000120	00	00	00	6F	99	99	99	20	99	00	00	4C	00	00	00	69	
00000130	00	00	00	76	00	99	00	65	00	00	00	64	00	00	00	2C	v e d ,
00000140	00	00	00	20	00	00	01	3E	00	00	01	40	00	00	00	43	
00000150	00																
00000160	00	00	99	20	99	99	00	4F	99	00	99	6E	99	99	00	65	
00000170	00																N U n d
00000180	00																
00000190	00																a b 1 @
000001A0	00																N o
00000180	00																1 N 5 7
000001C0	00																· · · · y · · ·] · · · · · · ·
000001D0	00																G 2 f f
000001E0	00																i] 0 :
000001F0	00						00									6C	h a 1
00000200	00	00	99	66	00	00	01	12	00	99	99	6C	00	00	00	6F	

-Indexação pelos atributos "casa" e "YearOfBirth"

```
FarquivoCesatet

1 Slytherin: [918, 1659, 3193, 3475, 3648, 3830, 4817, 4187, 4359, 4528, 4717, 4881, 5954, 5225, 5393, 12683, 13539, 13715, 14545, 14947, 15120, 17827, 17537, 17712, 17884, 18047, 18225, 18409, 188

2 Hufflepuff: [1283, 8525, 8683, 8841, 9010, 9193, 9358, 9514, 9675, 9852, 16807, 11471, 11628, 11805, 11805, 11805, 11805, 11805, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807, 11807
```

```
### arquivolanoitat

1 1954: [4187, 4359, 4528, 4717, 4881, 5954, 5225, 9675, 9852, 10907, 10206, 10414, 10582, 10752, 15312, 15487, 15649, 15812, 16602, 16186, 16357, 20725, 20897, 21075, 21262, 21417, 21589, 21751, 2

1 1954: [4187, 4359, 4528, 4717, 4881, 5954, 5225, 9675, 9852, 10907, 10206, 10414, 10582, 10752, 15312, 15487, 15649, 15812, 16602, 16186, 16357, 20725, 20897, 21075, 21262, 21417, 21589, 21751, 2

1 1952: [1308, 7838, 1286, 13804, 1938, 4921, 30224, 33366, 4917, 49574, 51244, 56342, 61084, 66815]

5 1969: [1659, 2825, 5573, 5756, 5753, 8525, 11118, 11298, 13964, 14856, 16892, 18859, 18567, 19543, 22113, 22271, 24028, 25907, 27693, 27858, 29457, 30392, 32742, 32901, 34599, 35561, 38064, 38215

6 1928: [1861, 7674, 13226, 18719, 24088, 29611, 34665, 40897, 45322, 59543, 55715, 66815]

7 1977: [1283, 7283, 12683, 18225, 23653, 29144, 34188, 39643, 44818, 59042, 55235, 66921, 66616]

8 1881: [5955, 6133, 1471], 11638, 17297, 172904, 22456, 22514, 28097, 28166, 33858, 3326, 33363, 3390, 33551, 34729, 43889, 48869, 49932, 54251, 5916, 69311, 64481, 64643]

9 1979: [465, 1453, 6554, 7352, 11980, 12865, 17537, 18409, 22979, 23809, 28533, 29299, 33551, 34350, 38909, 39742, 44204, 44098, 49397, 50202, 54574, 55391, 59637, 66484, 664944, 65777]

10 1980: [240, 712, 918, 2691, 5393, 6322, 6785, 6671, 7834, 16928, 11805, 12147, 12228, 13382, 16521, 17371, 17712, 17884, 18881, 21916, 22816, 23143, 23387, 24389, 27523, 28854, 2695, 2865]

1 1981: [2442, 2418, 7999, 3365, 13939, 33751, 98641, 16226, 11805, 12147, 12228, 13382, 16524, 17371, 17712, 17884, 18881, 21916, 22816, 23143, 23387, 24389, 27523, 28854, 2695, 28652, 2976

1 1981: [2242, 2418, 7999, 3365], 34921, 36861, 19226, 11805, 12147, 12226, 13382, 16254, 13495, 36861, 29143, 23387, 24399, 25272, 28864, 26652]

1 1981: [2242, 2418, 7999, 3365], 34229, 34681, 34661, 61852, 67190]

1 1981: [2342, 2418, 7418, 7499, 3365, 34224, 36814, 36814, 61852, 67190]

1 1981: [2342, 2418, 7419, 4395, 13969, 25272, 38644, 24292, 54598, 24762, 25586, 26214, 7378]
```

Conclusão

Os trabalhos práticos realizados ao longo deste semestre foram muito importantes para colocar em prática os conceitos que aprendemos em sala de aula. Conseguimos ver na prática o

impacto de escolher algoritmos eficazes e bem estruturados no desempenho e no funcionamento do programa.

Ao longo do semestre, aprendemos diferentes algoritmos de manipulação de arquivos indexados, compactação, casamento de padrões e criptografía e os trabalhos práticos foram importantes para ver na prática o funcionamento desses algoritmos.