Atividade Prática:
Índice linear

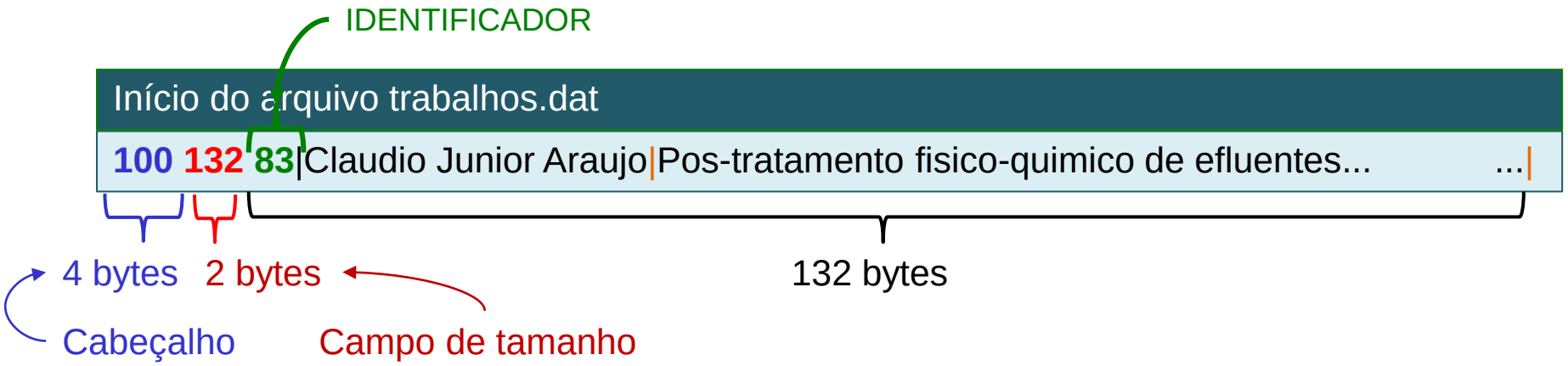
Organização e Recuperação de Dados

Profa. Valéria

UEM – CTC – DIN

Exercício

- Implementaremos um índice linear primário para o arquivo ***trabalhos.dat*** (disponível no *Classroom*)
- Formato do arquivo ***trabalhos.dat***
  - O **cabeçalho** é um inteiro de 4 bytes e armazena o total de registros
  - O primeiro campo de cada registro é um inteiro de 2 bytes que armazena o **tamanho do registro** em bytes
  - Outros campos: **IDENTIFICADOR**|AUTOR|TÍTULO|CURSO|TIPO
    - IDENTIFICADOR é um código que não se repete e será usada como chave primária

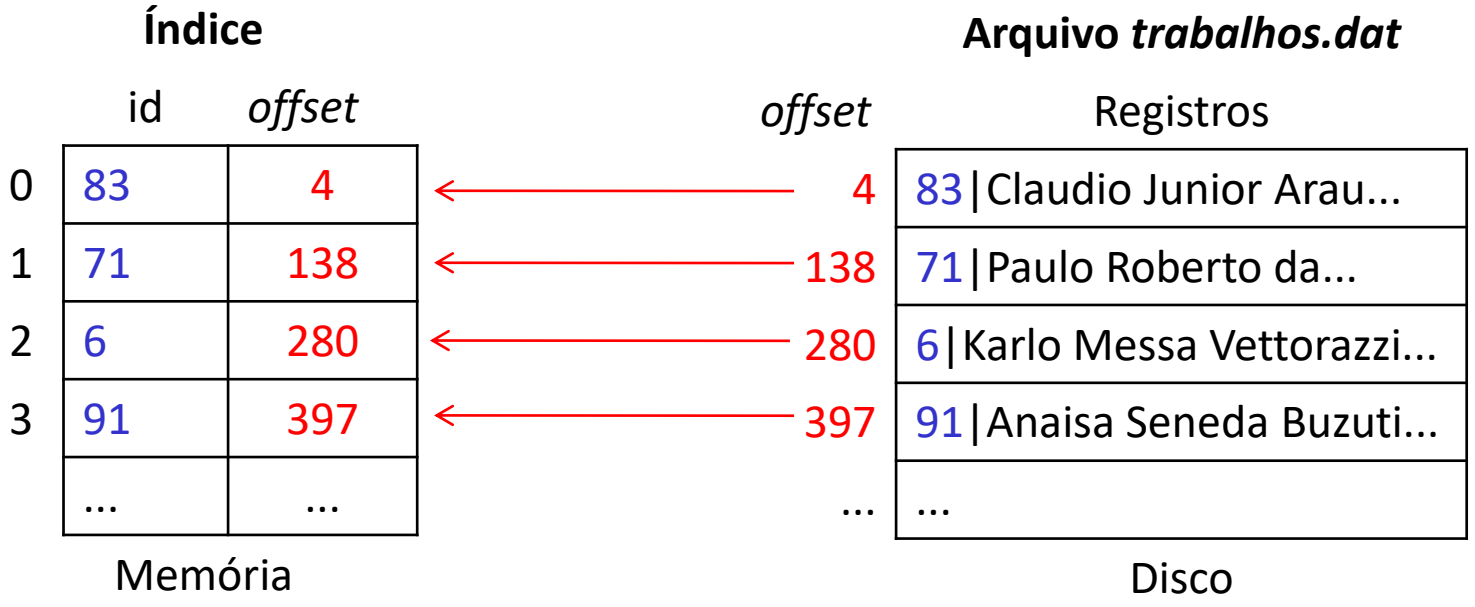


Exercício

- O programa deverá:
  - Construir um índice linear para o arquivo ***trabalhos.dat*** usando o **campo IDENTIFICADOR**
    - Use uma lista para armazenar o índice
    - Cada elemento da lista deverá ter dois campos: **chave** e ***offset***
  - Implementar uma busca por IDENTIFICADOR
    - Faça uma **busca binária** na lista **índice** e recupere o ***offset*** da **chave** buscada
    - Faça **acesso direto** no arquivo ***trabalhos.dat*** (*seek* + leitura) e imprima os dados do registro na tela

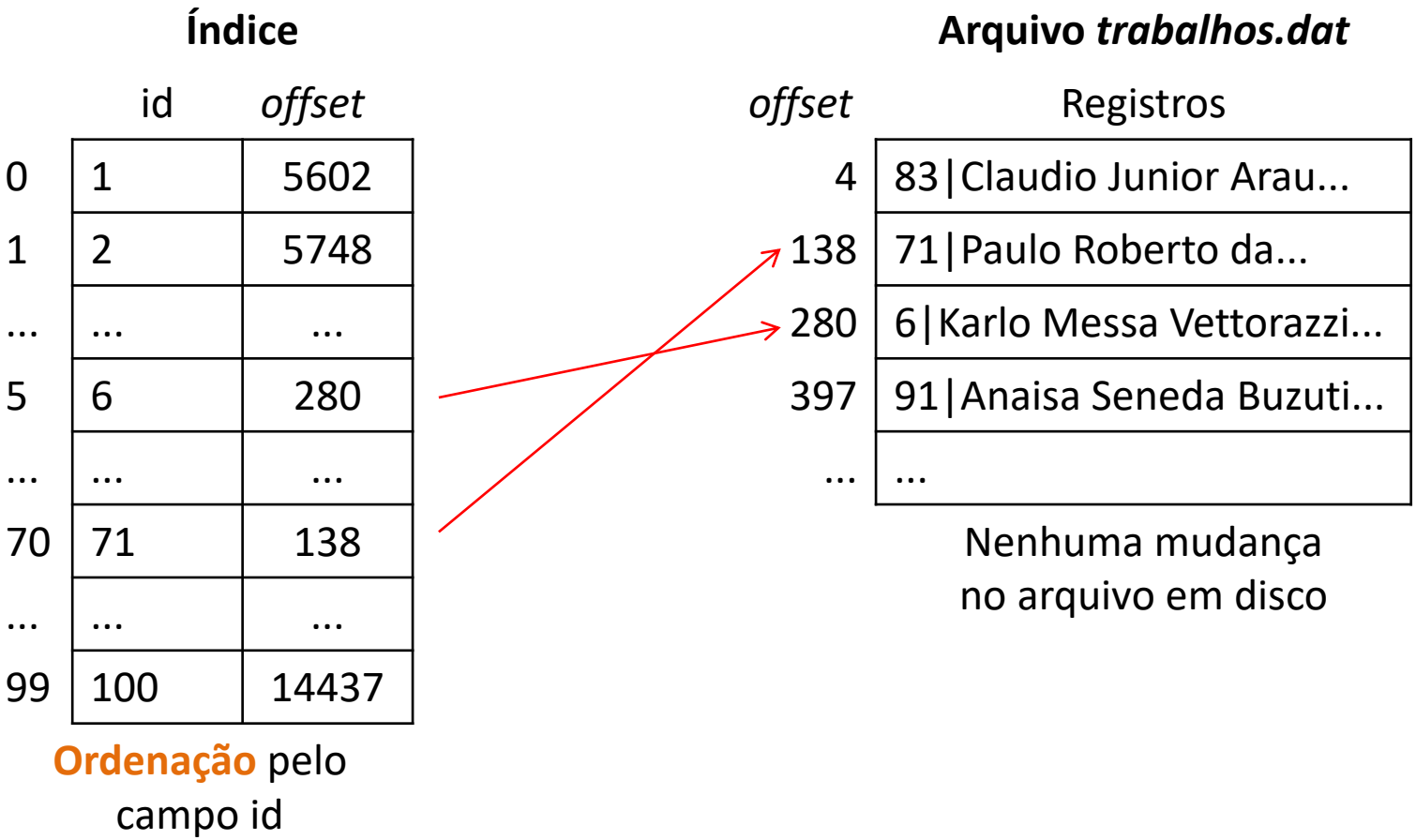
Construção do índice

- **Passo 1**: Leia o arquivo de registros sequencialmente e, para cada registro, insira a respectiva **chave** e ***byte-offset*** na lista **índice**



Construção do índice

- **Passo 2**: Ordene a lista **índice**



Construção do índice para o arquivo *trabalhos.dat*

Lista inicial:

	id	offset
0	**100**	4
1	**83**	129
2	**71**	263
...	**...**	...
99	**97**	14664

Índice:

	id	offset
0	**1**	5727
1	**2**	5873
3	**3**	6044
...	**...**	...
99	**100**	4

Pseudocódigo

```
from dataclasses import dataclass
# CONSTANTES GLOBAIS
SIZEOF_TOTALREG = 4
SIZEOF_TAMREG = 2

Declare um dataclass ELEMINDICE com dois campos do tipo int: ID e OFFSET
Abra o arquivo trabalhos.dat para leitura binária ("rb")

# construção do índice
Leia o cabeçalho do arquivo (SIZEOF_TOTALREG bytes) e armazene em TOTALREG
Crie a lista vazia INDICE
Faça OFFSET receber o offset do 1° registro
Para i até TOTALREG faça
    Leia um registro para REG e armazene o seu tamanho em TAMREG
    # você pode usar uma função similar à leia_reg() da Atividade 1
    Faça ID receber o 1° campo do registro
    # use split para separar os campos do registro e
    # não se esqueça de converter REGID para int
    Crie um ELEMINDICE com ID e OFFSET e insira-o em INDICE
    Faça OFFSET receber o offset do próximo registro
    # REGOFFSET + TAMREG + SIZEOF_TAMREG
Ordene o vetor INDICE pelo campo ID
# utilize qualquer algoritmo de ordenação de listas que você conheça


                                    # continua no próximo slide
```

Busca por ID usando o índice

Índice:

	id	offset
0	**1**	5727
1	**2**	5873
3	**3**	6044
...	**...**	...
99	**100**	4

Pseudocódigo

```
# início da busca
Receba o ID_BUSCADO da entrada padrão
Busque ID_BUSCADO em INDICE
# adapte a busca binária disponível nos slides da aula 6
# ela retorna o índice do elemento encontrado ou -1 caso contrário
Se ID_BUSCADO foi encontrado
    Posicione o ponteiro de L/E do arquivo em INDICE[i].OFFSET
    Leia o registro para REG
    Imprima o conteúdo do registro na tela
Senão imprima uma mensagem de erro
Feche o arquivo trabalhos.dat
# Fim PROGRAMA
```

Exemplo:

```
Digite o Id: 3

Id: 3
Nome: Andre Luiz Barcellos Junior
Título: Analise termoeconomica de coletores solares planos
Curso: Engenharia Mecânica
Tipo: Mestrado
```