

Trabalho final

Filipe da Silva de Oliveira
Ricardo Stoklosa

Projeto de Arquivos
Prof. Rafael Alceste Berri

Dados

- Nomes populares de bebes por sexo e etnia
- Cada registro possui o nome , ranking e ano
- O ranking varia por ano
- Department of Health and Mental Hygiene
- 2011 até 2016

Dados

Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
2016	MALE	WHITE NON HISPANIC	Graham	26	83
2016	FEMALE	BLACK NON HISPANIC	Lyric	18	35
2016	MALE	WHITE NON HISPANIC	Emmett	24	85
2016	FEMALE	WHITE NON HISPANIC	Naomi	53	44
2016	FEMALE	HISPANIC	Ruby	24	60
2016	MALE	HISPANIC	Harrison	12	89
2016	FEMALE	WHITE NON HISPANIC	Edith	12	82
2016	MALE	WHITE NON HISPANIC	Brayden	13	96
2016	MALE	HISPANIC	Abraham	42	62
2016	FEMALE	ASIAN AND PACIFIC ISLA...	Nina	18	33
2016	MALE	WHITE NON HISPANIC	Felix	23	86
2016	FEMALE	WHITE NON HISPANIC	Frieda	13	81
2016	FEMALE	BLACK NON HISPANIC	Ariana	23	30
2016	MALE	HISPANIC	Gavin	10	91

Índice

- Id
- Ano e ranking

Etapas

- Desenvolvemos o sistema seguindo as seguintes Etapas
 - Etapa 1
 - Geração do arquivo binário dado o CSV
 - Etapa 2
 - Manipulação deste arquivo Binário
 - Remoção e Inserção Lógica
 - Limpeza Física
 - Etapa 3
 - Criação de índices e buscas utilizando árvore B

Etapa 1 - Geração do Arquivo Binário

id
Ano de nascimento
Genero
Etnia
Nome
Numero de bebes
Rank

```
typedef struct Registro  
{  
    int id;  
    int anoNascimento;  
    char sexo[7];  
    char etnia[100];  
    char nome[100];  
    int contador;  
    int rank;  
}Registro;
```

Etapa 2 - Manipulação do Arquivo Binário

- Similar a Atividade 04
- Arquivo binário com um cabeçalho que indica o RRN desativado.
- Inserção leva o cabeçalho em consideração

```
typedef struct Cabecalho
{
    int topo;
}Cabecalho;
```

```
typedef struct Removido
{
    char character;
    int proximoPilha;
}Removido;
```

Etapa 3 - Buscas e Uso de Índices

- Árvore escolhida
 - Arvore B
 - A Árvore está toda em memória principal
 - Número máximo de chaves por nó = 5
 - Número mínimo de chaves por nó = 3
 - Número máximo de filhos por nó = 6
- Índices
 - Um arquivo para cada índice é gerado a partir do arquivo binário original
- Foi desenvolvido uma busca sequencial, em memória principal para a comparação do tempo de execução de cada algoritmo

Etapa 3 - Buscas e Uso de Índices

- id
 - Numérico
- Ano e ranking
 - Numérico e composto
 - O Ano e o Ranking são concatenados e string
 - Sofrem casting para inteiro novamente
 - Ex:
 - Ano 2011 + Ranking 1
 - 20111

```
typedef struct id
{
    long long id;
    long long rrn;
}id;

typedef struct anoRank
{
    int anoRank;
    long long rrn;
}anoRank;
```



Obrigado!

Filipe da Silva de Oliveira
Ricardo Stoklosa

Projeto de Arquivos
Prof. Rafael Alceste Berri