

# EP2 - mac323

---

Aluno: Davi de Menezes Pereira

NUSP: 11221988

---

## Introdução

O objetivo deste EP é fazer alguns experimentos sobre dicionários de palavras com grafos. Para isso criei o arquivo `interface.hpp` e o `ep2.cpp` que contém o código. No arquivo `interface.hpp` esta a implementação do grafo segundo as especificações do enunciado.

OBS: tentei mandar todos os arquivos `.txt` usados nos testes, mas não deu por causa do limite de espaço permitido no moodle.

## Como foram feitos os testes

### *Compilação*

```
g++ -Wall -g -O0 ep2.cpp -o ep2
```

### *Execução*

Iniciar o programa: `./ep2`, depois digitar o nome do arquivo `.txt` (com o `.txt`).

### *Código*

A implementação do grafo está em `interface.hpp`, e o programa principal está em `ep2.cpp`.

Foram usadas as bibliotecas `iostream`, `algorithm`, `vector`, `unordered_map`, `iterator`, `string` e `queue`.

As funções implementadas foram:

```
Grafo(int k);
int insere(std::string palavra);
int vertices();
int arestas();
std::string find(std::string p);
void uni(std::string p, std::string q);
int componentes();
bool conexo();
int tamComp(std::string palavra);
void bfs(std::string fonte);
int distancia(std::string a, std::string b);
void dfs1(std::string fonte, std::string a);
void dfs2(std::string fonte, std::string a, std::string b);
bool emCiclo(std::string a);
bool emCiclo(std::string a, std::string b);
double distanciaMedia()
double grauMedio();
```

```
bool denso();
double tamMedio();
int maiorComponente();
int menorComponente();
void imprime();
```

A descrição e implementação de cada uma delas está no código em interface.hpp.

Para cada teste, a primeira parte apresenta uma visão geral sobre o grafo, calculando, respectivamente:

- medidas

Quantidade de vertices, quantidade de arestas e grau medio dos vertices, densidade do grafo. (como a densidade do grafo não ficou muito bem especificada no enunciado, escolhi comparar a quantidade de arestas com  $v*v/100$ , sendo  $v$  a quantidade de vertices, como é possível ver no código).

- conexidade

Conexidade do grafo, quantas componentes, tamanho medio das componentes do grafo, tamanho da menor componente e tamanho da maior componente.

- caminhos

A distância media entre duas palavras.

Depois disso são feitos testes relacionados a ciclos no grafo. O usuario digita quantos testes deseja realizar e para cada teste digita duas palavras. O programa responde para cada teste se a primeira palavra está num ciclo sozinha, se a segunda palavra está num ciclo sozinha e se as duas palavras estão num mesmo ciclo.

## Testes

### 1) *dcc.txt*

Texto no site do Departamento de Ciência da Computação do IME-USP.

```
Entrada:
dcc.txt
2
nos aos
pela pelo

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 124
Quantidade de arestas: 17
Grau medio dos vertices: 0.274194
O grafo não eh considerado denso.
```

```
O grafo não eh conexo
Quantidade de componentes: 109
Tamanho medio das componentes do grafo: 1.13761
Tamanho da menor componente: 1
Tamanho da maior componente: 5
Distancia media entre duas palavras: 1.27273
```

```
Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    nos esta num cliclo.
  aos esta num cliclo.
  nos e aos estao num mesmo cliclo.
Digite as duas palavras:
  pela não esta num cliclo.
  pelo não esta num cliclo.
  pela e pelo não estao num mesmo cliclo.
```

## 2) *lusiadas.txt*

Livro de camões em português.

```
Entrada:
lusiadas.txt
2
a as
do de

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
  Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 8185
Quantidade de arestas: 8803
Grau medio dos vertices: 2.15101
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 3378
Tamanho medio das componentes do grafo: 2.42303
Tamanho da menor componente: 1
Tamanho da maior componente: 2996
Distancia media entre duas palavras: 7.72046

Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    a não esta num cliclo.
  as não esta num cliclo.
  a e as não estao num mesmo cliclo.
Digite as duas palavras:
  do não esta num cliclo.
```

```
de não esta num cliclo.  
do e de não estao num mesmo cliclo.
```

### 3) *lusiadas2.txt*

Livro de camões em inglês.

```
Entrada:  
lusiadas2.txt  
0  
  
Saida:  
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...  
  
Informações gerais:  
Imprimindo informações sobre o grafo...  
Quantidade de vertices: 7982  
Quantidade de arestas: 6177  
Grau medio dos vertices: 1.54773  
O grafo não eh considerado denso.  
O grafo não eh conexo  
Quantidade de componentes: 4399  
Tamanho medio das componentes do grafo: 1.8145  
Tamanho da menor componente: 1  
Tamanho da maior componente: 2045  
Distancia media entre duas palavras: 7.53131  
  
Digite quantos testes você deseja realizar:
```

### 4) *chinook.txt*

Livro retirado do site [gutenberg.org](http://gutenberg.org) em inglês.

```
Entrada:  
chinook.txt  
2  
had her  
to the  
  
Saida:  
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...  
  
Informações gerais:  
Imprimindo informações sobre o grafo...  
Quantidade de vertices: 3530  
Quantidade de arestas: 2363  
Grau medio dos vertices: 1.33881
```

```
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 2076
Tamanho medio das componentes do grafo: 1.70039
Tamanho da menor componente: 1
Tamanho da maior componente: 964
Distancia media entre duas palavras: 7.27926
```

```
Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    had esta num ciclo.
her esta num ciclo.
had e her estao num mesmo ciclo.
  Digite as duas palavras:
    to não esta num ciclo.
the esta num ciclo.
to e the não estao num mesmo ciclo.
```

## 5) receita.txt

Receita escolhida na internet.

```
Entrada:
receita.txt
2
de e
picados picadas

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
  Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 174
Quantidade de arestas: 18
Grau medio dos vertices: 0.206897
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 156
Tamanho medio das componentes do grafo: 1.11538
Tamanho da menor componente: 1
Tamanho da maior componente: 4
Distancia media entre duas palavras: 1.21739

Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    de não esta num ciclo.
e não esta num ciclo.
de e e não estao num mesmo ciclo.
  Digite as duas palavras:
```

```
picados não esta num cliclo.  
picadas esta num cliclo.  
picados e picadas não estao num mesmo cliclo.
```

## 6) *constituicao.txt*

Constituição brasileira.

```
Entrada:  
constituicao.txt  
2  
do no  
a e  
  
Saida:  
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...  
  
Informações gerais:  
Imprimindo informações sobre o grafo...  
Quantidade de vertices: 6006  
Quantidade de arestas: 3211  
Grau medio dos vertices: 1.06926  
O grafo não eh considerado denso.  
O grafo não eh conexo  
Quantidade de componentes: 3737  
Tamanho medio das componentes do grafo: 1.60717  
Tamanho da menor componente: 1  
Tamanho da maior componente: 846  
Distancia media entre duas palavras: 8.06629  
  
Digite quantos testes você deseja realizar:  
Digite as duas palavras:  
do não esta num cliclo.  
no não esta num cliclo.  
do e no não estao num mesmo cliclo.  
Digite as duas palavras:  
a não esta num cliclo.  
e não esta num cliclo.  
a e e não estao num mesmo cliclo.
```

## 7) *chocolate.txt*

Receita de chocolate quente.

```
Entrada:  
chocolate.txt  
2  
do no
```

a e

Saida:

Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...

Informações gerais:

Imprimindo informações sobre o grafo...

Quantidade de vertices: 56

Quantidade de arestas: 4

Grau medio dos vertices: 0.142857

O grafo não eh considerado denso.

O grafo não eh conexo

Quantidade de componentes: 52

Tamanho medio das componentes do grafo: 1.07692

Tamanho da menor componente: 1

Tamanho da maior componente: 3

Distancia media entre duas palavras: 1.2

Digite quantos testes você deseja realizar:

Digite as duas palavras:

do não esta num cliclo.

no não esta num cliclo.

do e no não estao num mesmo cliclo.

Digite as duas palavras:

a não esta num cliclo.

e não esta num cliclo.

a e e não estao num mesmo cliclo.

## 8) *pao.txt*

Receita de Pão em português.

Entrada:

pao.txt

2

um o

pao por

Saida:

Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...

Informações gerais:

Imprimindo informações sobre o grafo...

Quantidade de vertices: 88

Quantidade de arestas: 3

Grau medio dos vertices: 0.0681818

O grafo não eh considerado denso.

O grafo não eh conexo

Quantidade de componentes: 85

```
Tamanho medio das componentes do grafo: 1.03529
Tamanho da menor componente: 1
Tamanho da maior componente: 3
Distancia media entre duas palavras: 1.25
```

```
Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    um não esta num ciclo.
    o não esta num ciclo.
    um e o não estao num mesmo ciclo.
  Digite as duas palavras:
    pao não esta num ciclo.
    por não esta num ciclo.
    pao e por não estao num mesmo ciclo.
```

## 9) *espagete.txt*

Receita de espagete em português.

```
Entrada:
espagete.txt
2
um o
a e
```

```
Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...
```

```
Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 57
Quantidade de arestas: 4
Grau medio dos vertices: 0.140351
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 53
Tamanho medio das componentes do grafo: 1.07547
Tamanho da menor componente: 1
Tamanho da maior componente: 2
Distancia media entre duas palavras: 1
```

```
Digite quantos testes você deseja realizar:
  Digite as duas palavras:
    um não esta num ciclo.
    o não esta num ciclo.
    um e o não estao num mesmo ciclo.
  Digite as duas palavras:
    a não esta num ciclo.
    e não esta num ciclo.
    a e e não estao num mesmo ciclo.
```



10) *alice2.txt*

Alice no país das maravilha em português.

```
Entrada:
alice2.txt
0

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 3452
Quantidade de arestas: 2019
Grau medio dos vertices: 1.16976
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 2036
Tamanho medio das componentes do grafo: 1.69548
Tamanho da menor componente: 1
Tamanho da maior componente: 802
Distancia media entre duas palavras: 8.5879

Digite quantos testes você deseja realizar:
```

11) *alice.txt*

Livro Alice no Pais das Maravilha em inglês.

```
Entrada:
alice.txt
2
no in
cat cats

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 3008
Quantidade de arestas: 1853
Grau medio dos vertices: 1.23205
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 1845
```

```
Tamanho medio das componentes do grafo: 1.63035
Tamanho da menor componente: 1
Tamanho da maior componente: 775
Distancia media entre duas palavras: 7.4439
```

Digite quantos testes você deseja realizar:

Digite as duas palavras:

no não esta num cliclo.

in não esta num cliclo.

no e in não estao num mesmo cliclo.

Digite as duas palavras:

cat esta num cliclo.

cats esta num cliclo.

cat e cats estao num mesmo cliclo.

## 12) *tesouro.txt*

Livro A ilha do tesouro em inglês.

Entrada:

tesouro.txt

2

or of

this the

Saida:

Digite o nome do arquivo texto a ser lido (com o .txt se tiver):

Montando o grafo...

Informações gerais:

Imprimindo informações sobre o grafo...

Quantidade de vertices: 6764

Quantidade de arestas: 5446

Grau medio dos vertices: 1.61029

O grafo não eh considerado denso.

O grafo não eh conexo

Quantidade de componentes: 3764

Tamanho medio das componentes do grafo: 1.79702

Tamanho da menor componente: 1

Tamanho da maior componente: 2111

Distancia media entre duas palavras: 8.19751

Digite quantos testes você deseja realizar:

Digite as duas palavras:

or não esta num cliclo.

of não esta num cliclo.

or e of não estao num mesmo cliclo.

Digite as duas palavras:

this esta num cliclo.

the esta num cliclo.

this e the estao num mesmo cliclo.

### 13) *debian.txt*

Manual do Debian em inglês.

```
Entrada:
debian.txt
0

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 3049
Quantidade de arestas: 1058
Grau medio dos vertices: 0.693998
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 2195
Tamanho medio das componentes do grafo: 1.38907
Tamanho da menor componente: 1
Tamanho da maior componente: 394
Distancia media entre duas palavras: 11.8269

Digite quantos testes você deseja realizar:
```

### 14) *debiande.txt*

Manual do Debian em alemão.

```
Entrada:
debiande.txt
0

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 4222
Quantidade de arestas: 1668
Grau medio dos vertices: 0.790147
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 3048
Tamanho medio das componentes do grafo: 1.38517
Tamanho da menor componente: 1
```

```
Tamanho da maior componente: 493
Distancia media entre duas palavras: 8.67588

Digite quantos testes você deseja realizar:
```

### 15) *ubuntu.txt*

Manual do Ubuntu em inglês.

```
Entrada:
ubuntu.txt
0

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 7842
Quantidade de arestas: 3924
Grau medio dos vertices: 1.00077
O grafo não eh considerado denso.
O grafo não eh conexo
Quantidade de componentes: 5331
Tamanho medio das componentes do grafo: 1.47102
Tamanho da menor componente: 1
Tamanho da maior componente: 1335
Distancia media entre duas palavras: 7.37146

Digite quantos testes você deseja realizar:
```

### 16) *monteiro.txt*

Conto do Monteiro Lobato em português.

```
Entrada:
monteiro.txt
0

Saida:
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):
Montando o grafo...

Informações gerais:
Imprimindo informações sobre o grafo...
Quantidade de vertices: 1541
Quantidade de arestas: 613
Grau medio dos vertices: 0.795587
```

```
0 grafo não eh considerado denso.  
0 grafo não eh conexo  
Quantidade de componentes: 1081  
Tamanho medio das componentes do grafo: 1.42553  
Tamanho da menor componente: 1  
Tamanho da maior componente: 299  
Distancia media entre duas palavras: 8.87086
```

Digite quantos testes você deseja realizar:

### 17) *clarice.txt*

Texto da Clarice Lispector em português.

```
Entrada:  
clarice.txt  
0  
  
Saida:  
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...  
  
Informações gerais:  
Imprimindo informações sobre o grafo...  
Quantidade de vertices: 462  
Quantidade de arestas: 131  
Grau medio dos vertices: 0.5671  
0 grafo não eh considerado denso.  
0 grafo não eh conexo  
Quantidade de componentes: 352  
Tamanho medio das componentes do grafo: 1.3125  
Tamanho da menor componente: 1  
Tamanho da maior componente: 17  
Distancia media entre duas palavras: 2.13333  
  
Digite quantos testes você deseja realizar:
```

### 18) *fernando.txt*

Texto do Fernando Pessoa em português.

```
Entrada:  
fernando.txt  
0  
  
Saida:  
Digite o nome do arquivo texto a ser lido (com o .txt se tiver):  
Montando o grafo...
```

```
Informações gerais:  
Imprimindo informações sobre o grafo...  
Quantidade de vertices: 273  
Quantidade de arestas: 70  
Grau medio dos vertices: 0.512821  
O grafo não eh considerado denso.  
O grafo não eh conexo  
Quantidade de componentes: 212  
Tamanho medio das componentes do grafo: 1.28774  
Tamanho da menor componente: 1  
Tamanho da maior componente: 14  
Distancia media entre duas palavras: 2.34921
```

Digite quantos testes você deseja realizar:

## Conclusão

Dentre os textos escolhidos, podemos tentar observar alguns padrões:

Comparando os textos de receita com os livros no geral, os de receita parecem ser bem menos conexos, embora um fator importante talvez seja a diferença entre os tamanhos dos textos. Além disso, podemos ver que o primeiro texto (também pequeno, embora um pouco maior) também tem poucas arestas em comparação ao número de vértices, apesar de ser um pequeno texto de um site, e não um texto literário. Ainda olhando pras receitas podemos comparar com os últimos textos, que são literários, mas alguns deles são poemas, crônicas ou contos. Nesses últimos a diferença é menor, mas ainda são mais conexos (tem mais arestas em relação ao número de vértices) do que as receitas pequenas.

Comparando os textos de livros no geral com os mais técnicos, como os manuais e a constituição, podemos ver que os livros tendem a ser mais conexos.

Comparando os textos em inglês e seus respectivos em português, como nos livros "Os Lusíadas" e "Alice no País das Maravilhas", podemos ver q a lingua portuguesa aparenta ser mais conexa que a inglesa e comparando o manual do Debian em alemão e em inglês, elas parecem ter conexidade semelhantes.