

Trabalho Prático 3

Consultas ao Sistema de Despacho CabeAí

1 Descrição do problema

A empresa CabeAí ficou muito satisfeita com o sistema de despacho que você implementou e decidiu te contratar para implementar as consultas ao sistema logístico. Há basicamente uma consulta a ser implementada, que é a identificação dos destinos com base em nomes parciais do logradouro de origem e destino.

Assim, a partir de uma base de dados de endereços, a entrada é a sua posição atual (coordenadas GPS) e uma parte do nome (uma ou mais palavras) do logradouro de destino. A resposta deve ser os n logradouros mais próximos em termos de distância que contenham as palavras informadas.

Apenas para ilustrar o cenário, vamos analisar um caso real, Belo Horizonte. A base de endereços de Belo Horizonte contém 725917 endereços diferentes, há 12134 nomes únicos de ruas. Esses 12134 nomes únicos de ruas contém 30105 palavras no total, ou seja, 2.48 palavras por nome de rua. Avaliando essas palavras que compõem os nomes de rua, encontramos 6911 palavras únicas. Entretanto, essa distribuição de palavras por nomes de rua é muito enviesada. Por exemplo, 4793 palavras ocorrem apenas 1 vez e apenas 422 palavras ocorrem 10 vezes ou mais e representam quase $\frac{2}{3}$ das ocorrências de palavras nos nomes de ruas. O enviesamento destes dados pode ser visualizado nos gráficos das Figuras 1 e 2, que apresentam, respectivamente, a distribuição de frequência e a distribuição de frequência cumulativa das palavras em nomes de ruas. Essas características da base de dados são um desafio do ponto de vista de implementação, pois as estruturas de dados têm que estar preparadas para garantir a eficiência computacional a despeito da popularidade das palavras e/ou dos logradouros.

A base de dados a ser utilizada pelo sistema a ser implementado é uma base de endereços, no caso específico da cidade de Belo Horizonte, que contém os seguintes atributos:

IdEnd: Identificador alfanumérico de endereço, isto é, associado à tupla logradouro, número, bairro e região.

IdLog: Identificador numérico do logradouro, isto é, associado à tupla logradouro, bairro e região.

TipoLog: Tipo do logradouro, por exemplo, rua, avenida, alameda ou praça, entre outros.

Log: Nome do logradouro, composto de uma ou mais palavras separadas por espaço em branco.

Num: Número do endereço no logradouro.

Bairro: Bairro do endereço (e não do logradouro), uma vez que um logradouro pode atravessar vários bairros.

Regiao: Região do endereço, que, no caso de Belo Horizonte, são nove: BARREIRO, CENTRO-SUL, LESTE, NORDESTE, NOROESTE, NORTE, OESTE, PAMPULHA e VENDA NOVA.

CEP: Código de Endereçamento Postal do endereço, conforme definido pelos Correios.

Lat: Latitude do endereço, convertida para um número real.

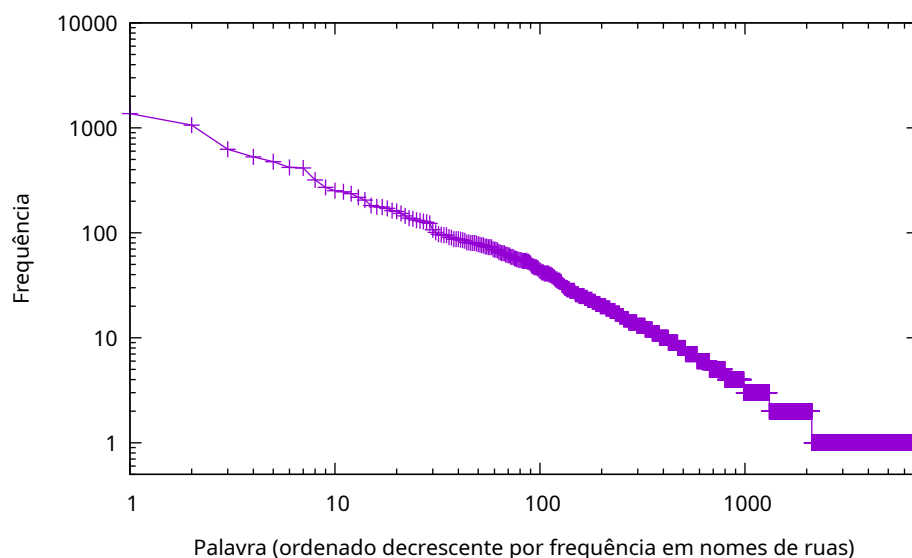


Figura 1: Distribuição de frequência de palavras por nomes de ruas

Long: Longitude do endereço, convertido para um número real.

Como mencionado, a tarefa a ser realizada é, a partir de uma consulta, identificar os logradouros que a satisfaçam em termos das suas palavras (ou seja, contenham as palavras da consulta, em qualquer ordem) e sejam mais próximos em termos de coordenadas latitude/longitude. A Figura 3 apresenta o formato da entrada do sistema de consultas, iniciando com a base de endereços (que contem os atributos descritos) a ser considerada (N endereços), seguida de M consultas, para as quais devem ser geradas até R respostas. Cada consulta possui um identificador numérico, a consulta propriamente dita, que consiste de um conjunto pelo menos unitário de palavras separadas por espaços, e as coordenadas latitude/longitude a partir de onde a consulta está sendo realizada. Um exemplo de entrada que segue esse formato é apresentada na Figura 4.

A saída contem, para cada uma das M consultas, o seu identificador, o número de respostas encontradas e cada uma das respostas, uma por linha. O formato da saída é apresentado na Figura 5 e a Figura 6 contem a saída para o arquivo de entrada apresentado na Figura 4

2 Arquitetura

Nesta seção vamos detalhar o sistema a ser implementado.

2.1 Descrição da Solução

Já apresentamos um exemplo de entrada e um exemplo de saída para o sistema a ser implementado. O funcionamento do sistema pode ser dividido em dois grandes estágios: (i) ativação do sistema de consulta e (ii) realização de consultas. Esses estágios são detalhados a seguir.

A ativação do sistema de consulta compreende a construção dos TADs e sua instanciação com a base de dados da entrada. Distinguimos as seguintes etapas nesse processo:

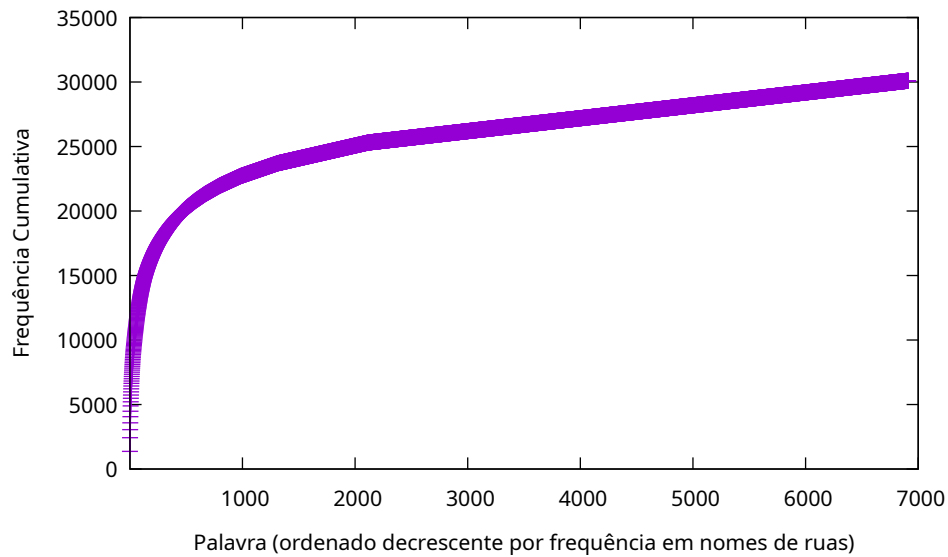


Figura 2: Distribuição cumulativa de frequência de palavras por nomes de ruas

```

N %NumEnderecos
IdEnd1;IdLog1;TipoLog1;Log1;Num1;Bairro1;Regiao1;CEP1;Lat1;Long1
...
IdEndN;IdLogN;TipoLogN;LogN;NumN;BairroN;RegiaoN;CEPN;LatN;LongN
M R %NumConsultas, MaximoRespostas
IdConsulta1;Consulta1;LatOrigem1;LongOrigem1
...
IdConsultaM;ConsultaM;LatOrigemM;LongOrigemM
    
```

Figura 3: Formato da Entrada

```

55
11446500063;114465;RUA;ADEMAR PIMENTA BRANT;63;Mangabeiras;CENTRO-SUL;30210050;613481.472577573;7792733.02007114
11446500087;114465;RUA;ADEMAR PIMENTA BRANT;87;Mangabeiras;CENTRO-SUL;30210050;613503.634441812;7792718.55572104
11446500180;114465;RUA;ADEMAR PIMENTA BRANT;180;Mangabeiras;CENTRO-SUL;30210050;613545.881207012;7792671.78219912
02014400120;20144;RUA;ARISTOTELES BRANDAO;120;Mantiqueira;VENDA NOVA;31660510;606517.334616207;7810355.71572825
02014400399;20144;RUA;ARISTOTELES BRANDAO;399;Mantiqueira;VENDA NOVA;31660510;606302.983893734;7810302.51655679
02014400421;20144;RUA;ARISTOTELES BRANDAO;421;Mantiqueira;VENDA NOVA;31660510;606286.22009228;7810317.87343094
00597100425;5971;RUA;ARISTOTELES CALDEIRA;425;Prado;OESTE;30411225;608188.527608476;7795837.15559551
00597100426;5971;RUA;ARISTOTELES CALDEIRA;426;Prado;OESTE;30411225;608164.433121334;7795840.35496482
00597100432;5971;RUA;ARISTOTELES CALDEIRA;432;Prado;OESTE;30411225;608165.488594266;7795834.31616155
00597100603;5971;RUA;ARISTOTELES CALDEIRA;603;Barroca;OESTE;30431054;608229.113851304;7795676.96733901
10459300032;104593;RUA;BRANT HORTA;32;Mantiqueira;VENDA NOVA;31660350;605804.6761582;7810772.49589893
10459300180;104593;RUA;BRANT HORTA;180;Maria Helena;VENDA NOVA;31660350;6055648.886575768;7810780.40435319
10459300205;104593;RUA;BRANT HORTA;205;Maria Helena;VENDA NOVA;31660350;605596.005685591;7810757.8188365
10459300215;104593;RUA;BRANT HORTA;215;Maria Helena;VENDA NOVA;31660350;605579.84374105;7810744.61145617
01164000010;11640;RUA;CALDEIRA BRANT;10;Sagrada Família;LESTE;31030180;612357.360808444;7798133.89978311
01164000041;11640;RUA;CALDEIRA BRANT;41;Sagrada Família;LESTE;31030180;612344.315819465;7798184.32978998
01164000048;11640;RUA;CALDEIRA BRANT;48;Sagrada Família;LESTE;31030180;612368.402059797;7798174.48173829
01164000054;11640;RUA;CALDEIRA BRANT;54;Sagrada Família;LESTE;31030180;612371.875820502;7798186.83266912
01164000061;11640;RUA;CALDEIRA BRANT;61;Sagrada Família;LESTE;31030180;612342.124814282;7798207.74841178
10982000020;109820;RUA;CONEGO CALDEIRA BRANT;20;Bonsucesso;BARREIRO;30622090;605402.356288039;7790077.00340136
10982000020A;109820;RUA;CONEGO CALDEIRA BRANT;20;Bonsucesso;BARREIRO;30622090;605393.038483985;7790067.695296
10982000020A;109820;RUA;CONEGO CALDEIRA BRANT;20;Bonsucesso;BARREIRO;30622090;605393.038483985;7790067.695296
12115200032;121152;RUA;EDUARDO BRANT RIBEIRO;32;Barreiro;BARREIRO;30640125;602600.345641822;7790433.62736377
12115200037;121152;RUA;EDUARDO BRANT RIBEIRO;37;Barreiro;BARREIRO;30640125;602626.904048147;7790439.51961224
12115200044;121152;RUA;EDUARDO BRANT RIBEIRO;44;Barreiro;BARREIRO;30640125;602604.386125407;7790418.88028619
12115200047;121152;RUA;EDUARDO BRANT RIBEIRO;47;Barreiro;BARREIRO;30640125;602629.435522346;7790427.31209704
08013700046;80137;RUA;HERBERT BRANT ALEIXO;46;Itapoã;PAMPULHA;31710300;608891.895130345;7805915.058127
08013700058;80137;RUA;HERBERT BRANT ALEIXO;58;Itapoã;PAMPULHA;31710300;608904.074307982;7805911.23888218
08013700082;80137;RUA;HERBERT BRANT ALEIXO;82;Itapoã;PAMPULHA;31710300;608926.247503015;7805906.15988561
08013700094;80137;RUA;HERBERT BRANT ALEIXO;94;Itapoã;PAMPULHA;31710300;608937.956665072;7805903.01050811
12064100044;120641;RUA;JOAO CALDEIRA FILHO;44;Mantiqueira;VENDA NOVA;31655550;606036.684352195;7811633.77558287
12064100044A;120641;RUA;JOAO CALDEIRA FILHO;44;Mantiqueira;VENDA NOVA;31655550;606050.388945415;7811635.9051494
12064100057;120641;RUA;JOAO CALDEIRA FILHO;57;Mantiqueira;VENDA NOVA;31655550;606017.53750174;7811654.97126855
02202300014;22023;RUA;JOAO EDMUNDO CALDEIRA BRANT;14;Itapoã;PAMPULHA;31710200;608641.802186465;7805989.77440295
02202300018;22023;RUA;JOAO EDMUNDO CALDEIRA BRANT;18;Itapoã;PAMPULHA;31710200;608643.888385072;7805994.46344851
02202300466;22023;RUA;JOAO EDMUNDO CALDEIRA BRANT;466;Itapoã;PAMPULHA;31710200;608728.956435839;7806433.86532994
08627500002;86275;RUA;JOSE ANTONIO ULHOA;2;Frei Leopoldo;NORTE;31746020;611210.334559607;7809456.03602834
08627500010;86275;RUA;JOSE ANTONIO ULHOA;10;Frei Leopoldo;NORTE;31746020;611223.907294106;7809462.88466908
08627500017;86275;RUA;JOSE ANTONIO ULHOA;17;Frei Leopoldo;NORTE;31746020;611215.686144691;7809476.92188836
08627500019;86275;RUA;JOSE ANTONIO ULHOA;19;Frei Leopoldo;NORTE;31746020;611223.041478994;7809485.18025064
12412200005;124122;RUA;MIGUEL CALDEIRA DOS SANTOS;5;Jardim dos Comerciantes;VENDA NOVA;31652122;607187.281025997;7811798.04246431
12412200006;124122;RUA;MIGUEL CALDEIRA DOS SANTOS;6;Jardim dos Comerciantes;VENDA NOVA;31652122;607189.523902513;7811760.02999725
12412200015;124122;RUA;MIGUEL CALDEIRA DOS SANTOS;15;Jardim dos Comerciantes;VENDA NOVA;31652122;607301.173267579;7811944.90334367
00013000015;130;RUA;OSCAR COELHO DOS SANTOS;15;Salgado Filho;OESTE;30550590;606929.949051997;7793870.21562208
00013000025;130;RUA;OSCAR COELHO DOS SANTOS;25;Salgado Filho;OESTE;30550590;606928.151446414;7793857.43815485
00013000028;130;RUA;OSCAR COELHO DOS SANTOS;28;Salgado Filho;OESTE;30550590;606905.747357172;7793793.63080475
01777600004;17776;RUA;OSCAR NEGRAO DE LIMA;4;Vila Nova Gameleira I;OESTE;30510210;605013.028749605;7794721.00812766
09087500030;90875;RUA;OSCAR VERSIANI CALDEIRA;30;Mangabeiras;CENTRO-SUL;30210280;613516.666152952;7793548.34846536
09087500056;90875;RUA;OSCAR VERSIANI CALDEIRA;56;Mangabeiras;CENTRO-SUL;30210280;613533.776376581;7793575.53307475
09087500068;90875;RUA;OSCAR VERSIANI CALDEIRA;68;Mangabeiras;CENTRO-SUL;30210280;613542.385087939;7793597.62869406
06519500011;65195;RUA;SANTO ANTONIO DE LISBOA;11;São João Batista;VENDA NOVA;31520250;608036.235627214;7808480.02751915
06519500011A;65195;RUA;SANTO ANTONIO DE LISBOA;11;São João Batista;VENDA NOVA;31520250;608032.772372609;7808467.41008739
30388400121;303884;PCA;VINICIUS CALDEIRA BRANT;121;Buritis;OESTE;30000000;608357.43565686;7790618.49992465
30388400121;303884;PCA;VINICIUS CALDEIRA BRANT;121;Buritis;OESTE;30000000;608357.43565686;7790618.49992465
30388400134;303884;PCA;VINICIUS CALDEIRA BRANT;134;Buritis;OESTE;30000000;608330.587047225;7790651.55337705
3 3
1;CALDEIRA BRANT;606526.987402;7808523.765932
2;BRANT HORTA;606526.987402;7808523.765932
3;BRANT VERSIANI;606500.843402;7808633.468966

```

Figura 4: Exemplo de Entrada

```

M %NumConsultas
IdConsulta1;R1 %NumRespostas
IdLog1;Logradouro1
...
IdLogR1;LogradouroR1
....
IdConsultaM;RM %NumRespostas
IdLog1;Logradouro1
...
IdLogRM;LogradouroRM

```

Figura 5: Formato da Saída

3
1;3
22023;JOAO EDMUNDO CALDEIRA BRANT
11640;CALDEIRA BRANT
303884;VINICIUS CALDEIRA BRANT
2;1
104593;BRANT HORTA
3;0

Figura 6: Exemplo de Saída

1. **Leia e armazene a base de endereços**, o que significa que deve ser projeto e implementado um tipo abstrato de dados para armazenar a base de endereços. Esse TAD deve ser compatível com as outras ações e garantir a eficiência computacional do sistema a ser implementado.
2. **Identifique os logradouros únicos**, que é uma ação necessária tendo em vista que a saída é apenas em termos de logradouros. Para tal, é necessário projetar e implementar um TAD logradouro, que vai naturalmente tratar as várias ocorrências de um dado logradouro na base de endereços. Importante lembrar que o número de endereços por logradouro é altamente variável, e a estrutura de dados escolhida deve ser capaz de lidar com tal variabilidade. A Tabela 1 identifica os logradouros únicos da base de endereços apresentada na Figura 4.
3. Calcule o centro de gravidade de cada logradouro único. A primeira versão do sistema a ser implementado não leva em consideração endereços individuais para a sugestão, mas o centro de gravidade dos endereços de cada logradouro. Desta forma, é necessário calcular as coordenadas médias de cada logradouro, que vão ser utilizadas para gerar a resposta da consulta. A Tabela 2 contém, para os logradouros únicos da entrada exemplo, a quantidade de endereços para cada logradouro e o seu centro de gravidade.
4. **Construa a estrutura de dados dos logradouros únicos**, incluindo as suas coordenadas médias, ou seja, entre os atributos do logradouro, o seu TAD deve armazenar o centro de gravidade, composto de uma latitude e uma longitude médias de todos os endereços do logradouro. Como mencionado, a Tabela 2 apresenta alguns dos atributos que devem constar do TAD logradouro.
5. **Identifique as palavras únicas dos logradouros únicos**. As palavras únicas vão ser os átomos da consulta e elas são determinadas pela decomposição do atributo logradouro em palavras e seu posterior armazenamento em um índice de palavras. Como as consultas sempre são compostas de palavras inteiras, o índice de palavras seria suficiente. As palavras únicas da base de endereço exemplo são apresentadas na Figura 7.
6. **Construa um índice com as palavras únicas**, onde cada palavra aponta para uma lista de logradouros que a contem. A Tabela 3 apresenta as palavras únicas com os conteúdos associados a cada uma.

Uma vez que o sistema de consulta esteja pronto e operacional, realizamos, para cada consulta, as seguintes etapas:

1. **Recupere as listas de logradouros das palavras da consulta.** Para a consulta CALDEIRA BRANT, temos duas listas de logradouros, apresentados na Tabela 4.
2. **Faça a junção das listas de palavras,** identificando os logradouros que contem todas as palavras da consulta. No caso da consulta CALDEIRA BRANT, temos quatro logradouros que contem ambas as palavras, os quais estão destacados na Tabela 4.
3. **Calcule a distância do ponto de partida até as coordenadas médias dos os logradouros,** isto é, para cada logradouro resultante na junção entre as listas das palavras da consulta, calcule a distância do ponto de origem até a sua coordenada média. A Tabela 5 apresenta os logradouros, as coordenadas médias e as distâncias até à origem para a consulta CALDEIRA BRANT.
4. **Ordene pela distância até a origem.** Para a consulta CALDEIRA BRANT, os logradouros da Tabela 5 estão ordenados.
5. **Retorne os logradouros ordenados até o número máximo especificado na consulta.** No caso do exemplo apresentado na Figura 4, as respostas às consultas estão limitadas a três logradouros, como pode ser visto no exemplo de saída (Figura 6). Nesse exemplo de saída também podemos verificar o comportamento esperado do sistema de consultas quando nenhum logradouro satisfaz a consulta, como é o caso da consulta 3.

2.2 Estruturas de Dados

A implementação do sistema de consulta deve demandar quatro tipos abstratos de dados, descritos a seguir.

2.2.1 TAD Endereço

O TAD Endereço basicamente armazena os dados de entrada. Uma função que nos parece pertencer a esse TAD é o cálculo das coordenadas médias, necessário para construir o TAD Logradouro.

2.2.2 TAD Logradouro

O TAD Logradouro vai conter as informações para os logradouros, em particular as coordenadas médias dos mesmos. Ele também fará a associação entre identificadores e nomes dos logradouros.

2.2.3 TAD Palavra

O TAD Palavra é o mais fundamental para as consultas. Entre outras funções, ele vai permitir retornar uma lista dos logradouros onde uma palavra ocorre. Há diversos compromissos e desafios na sua implementação, desde a busca eficiente por uma palavra quanto a manipulação das listas de logradouros. Pelo menos uma versão do índice de palavras deve ser implementado como uma árvore de pesquisa balanceada.

2.3 TAD Consulta

O TAD consulta contém as informações necessárias para satisfazer uma consulta. Ele vai obter informações dos TADs Logradouro e Palavra para poder gerar a resposta da consulta.

IdLog	Logradouro
114465	ADEMAR PIMENTA BRANT
20144	ARISTOTELES BRANDAO
5971	ARISTOTELES CALDEIRA
104593	BRANT HORTA
11640	CALDEIRA BRANT
109820	CONEGO CALDEIRA BRANT
121152	EDUARDO BRANT RIBEIRO
80137	HERBERT BRANT ALEIXO
120641	JOAO CALDEIRA FILHO
22023	JOAO EDMUNDO CALDEIRA BRANT
86275	JOSE ANTONIO ULHOA
124122	MIGUEL CALDEIRA DOS SANTOS
130	OSCAR COELHO DOS SANTOS
17776	OSCAR NEGRAO DE LIMA
90875	OSCAR VERSIANI CALDEIRA
65195	SANTO ANTONIO DE LISBOA
303884	VINICIUS CALDEIRA BRANT

Tabela 1: Logradouros únicos (e seus identificadores) para a amostra de base de endereços

2.4 Detalhes de implementação

Idealmente, considerando o custo de acessar a memória secundária, o arquivo de entrada deve ser lido uma vez só. Entretanto, a decisão de projeto se os vários TADs serão construídos enquanto o arquivo de entrada é lido ou não é sua, ou seja, você pode ir populando os TADs endereço, logradouro e palavra à medida que cada endereço for lido, ou pode ler todos os endereços, armazená-los no TAD endereço, e então popular os TADs logradouro e palavra posteriormente. Seja lá qual for a sua decisão é importante defendê-la na documentação a ser entregue, por exemplo em termos de eficiência computacional e localidade de referência temporal e espacial.

3 Análise Experimental

Considerando a natureza das consultas a serem implementadas, sugerimos que a análise experimental avalie o impacto de variações nas seguintes dimensões:

1. Número de endereços
2. Número de logradouros
3. Tamanho e natureza (isto é, popularidade dos palavras) da consulta.

Em termos de experimentação, sugerimos que seja desmembrado o tempo de leitura da base de endereços e construção dos TADs do tempo de processamento das consultas. Desta forma, as medições não serão dominadas pelos custos de leitura da memória secundária.

4 Pontos Extra

Os pontos extra tem por objetivo avaliar melhorias no sistema ou no processamento das consultas:

IdLog	Logradouro	Quant	Lat. Média	Long. Média
114465	ADEMAR PIMENTA BRANT	3	613510.329409	7792707.785997
20144	ARISTOTELES BRANDAO	3	606368.846201	7810325.368572
5971	ARISTOTELES CALDEIRA	4	608186.890794	7795797.198515
104593	BRANT HORTA	4	605657.353040	7810763.832636
11640	CALDEIRA BRANT	5	612356.815864	7798177.458478
109820	CONEGO CALDEIRA BRANT	3	605396.144419	7790070.797998
121152	EDUARDO BRANT RIBEIRO	4	602615.267834	7790429.834840
80137	HERBERT BRANT ALEIXO	4	608915.043402	7805908.866851
120641	JOAO CALDEIRA FILHO	3	606034.870266	7811641.550667
22023	JOAO EDMUNDO CALDEIRA BRANT	3	608671.549002	7806139.367727
86275	JOSE ANTONIO ULHOA	4	611218.242369	7809470.255709
124122	MIGUEL CALDEIRA DOS SANTOS	3	607225.992732	7811834.325268
130	OSCAR COELHO DOS SANTOS	3	606921.282619	7793840.428194
17776	OSCAR NEGRAO DE LIMA	1	605013.028750	7794721.008128
90875	OSCAR VERSIANI CALDEIRA	3	613530.942539	7793573.836745
65195	SANTO ANTONIO DE LISBOA	2	608034.504000	7808473.718803
303884	VINICIUS CALDEIRA BRANT	3	608348.486120	7790629.517742

Tabela 2: Logradouros únicos e suas coordenadas médias

ADEMAR
 ALEIXO
 ANTONIO
 ARISTOTELES
 BRANDAO
 BRANT
 CALDEIRA
 COELHO
 CONEGO
 DE
 DOS
 EDMUNDO
 EDUARDO
 FILHO
 HERBERT
 HORTA
 JOAO
 JOSE
 LIMA
 LISBOA
 MIGUEL
 NEGRAO
 OSCAR
 PIMENTA
 RIBEIRO
 SANTO
 SANTOS
 ULHOA
 VERSIANI
 VINICIUS

Figura 7: Palavras únicas dos logradouros únicos

Palavra	Quant.	Logradouros
ADEMAR	1	114465
ALEIXO	1	80137
ANTONIO	2	65195 86275
ARISTOTELES	2	5971 20144
BRANDAO	1	20144
BRANT	8	11640 22023 80137 104593 109820 114465 121152 303884
CALDEIRA	8	5971 11640 22023 90875 109820 120641 124122 303884
COELHO	1	130
CONEGO	1	109820
DE	2	17776 65195
DOS	2	130 124122
EDMUNDO	1	22023
EDUARDO	1	121152
FILHO	1	120641
HERBERT	1	80137
HORTA	1	104593
JOAO	2	22023 120641
JOSE	1	86275
LIMA	1	17776
LISBOA	1	65195
MIGUEL	1	124122
NEGRAO	1	17776
OSCAR	3	130 17776 90875
PIMENTA	1	114465
RIBEIRO	1	121152
SANTO	1	65195
SANTOS	2	130 124122
ULHOA	1	86275
VERSIANI	1	90875

Tabela 3: Conteúdo do Índice das Palavras Únicas

Palavra	Quant.	Logradouros
BRANT	8	11640 22023 80137 104593 109820 114465 121152 303884
CALDEIRA	8	5971 11640 22023 90875 109820 120641 124122 303884

Tabela 4: Logradouros da consulta, com a junção das duas listas em destaque

IdLog	Logradouro	Lat. Média	Long. Média	Dist. Origem
22023	JOAO EDMUNDO CALDEIRA BRANT	608671.549002	7806139.367727	3206.945471
11640	CALDEIRA BRANT	612356.815864	7798177.458478	11875.730623
303884	VINICIUS CALDEIRA BRANT	608348.486120	7790629.517742	17986.716650
109820	CONEGO CALDEIRA BRANT	605396.144419	7790070.797998	18487.585873

Tabela 5: Logradouros da junção, ordenados pela distância à origem (606526.987402, 7808523.765932)

- Realize uma avaliação comparativa de possíveis implementações do TAD palavra, ou seja, implemente as mesmas funções de formas diferentes e avalie comparativamente o desempenho dessas comparações como função do número de palavras e/ou número de logradouros.
- Estenda a consulta para considerar partes de palavras, por exemplo prefixos.
- Coordenadas médias podem fazer pouco sentido para logradouros muito longos ou que apresentem um trajeto não linear. Um caso típico em Belo Horizonte é a Avenida do Contorno, que tem um formato semi-circular suas coordenadas médias devem mapear para um lugar próximo ao Palácio das Artes. Para evitar esse tipo de problema, você pode segmentar esses logradouros, de tal forma que nenhum logradouro ocupe uma área maior que uma distância pré-definida.

Esteja à vontade para propor aperfeiçoamentos ao sistema de consultas.

5 Como será feita a entrega

5.1 Submissão

A entrega do TP3 compreende duas submissões:

VPL TP3: Submissão do código a ser submetido até **08/12/25, 7:59**. **Não serão aceitas submissões em atraso.** Detalhes sobre a submissão do código são apresentados na Seção 5.3.

Relatório TP3: Arquivo PDF contendo a documentação do TP, assim como a avaliação experimental, conforme instruções, a ser submetido até **08/12/25, 7:59**. **Não serão aceitas submissões em atraso.** Detalhes sobre a submissão de relatório são apresentados na Seção 5.2.

5.2 Documentação

A documentação do trabalho deve ser entregue em formato **PDF** e também **DEVE** seguir o modelo de relatório que será postado no `minha.ufmg`. Além disso, a documentação deve conter **TODOS** os itens descritos a seguir **NA ORDEM** em que são apresentados:

1. **Capa:** Título, nome, e matrícula.
2. **Introdução:** Contém a apresentação do contexto, problema, e qual solução será empregada.
3. **Método:** Descrição da implementação, detalhando as estruturas de dados, tipos abstratos de dados (ou classes) e funções (ou métodos) implementados. A justificativa das decisões de projeto deve incluir argumentos em termos de eficiência computacional e maximização da localidade de referência, ambos diante das características da base de dados e da carga de trabalho.
4. **Análise de Complexidade:** Contém a análise da complexidade de tempo e espaço dos procedimentos implementados, formalizada pela notação assintótica.
5. **Estratégias de Robustez:** Contém a descrição, justificativa e implementação dos mecanismos de programação defensiva e tolerância a falhas implementados.

6. **Análise Experimental:** Apresenta os experimentos realizados em termos de desempenho computacional¹, assim como as análises dos resultados.
7. **Conclusões:** A Conclusão deve conter uma frase inicial sobre o que foi feito no trabalho. Posteriormente deve-se sumarizar o que foi aprendido.
8. **Bibliografia:** Contém fontes utilizadas para realização do trabalho. A citação deve estar em formato científico apropriado que deve ser escolhido por você.
9. Número máximo de páginas incluindo a capa: 10

A documentação deve conter a descrição do seu trabalho em termos funcionais, dando foco nos algoritmos, estruturas de dados e decisões de implementação importantes durante o desenvolvimento.

Evite a descrição literal do código-fonte na documentação do trabalho.

Dica: Sua documentação deve ser clara o suficiente para que uma pessoa (da área de Computação ou não) consiga ler, entender o problema tratado e como foi feita a solução.

A documentação deverá ser entregue como uma atividade separada designada para tal no minha.ufmg. A entrega deve ser um arquivo `.pdf`, nomeado `nome_sobrenome_matricula.pdf`, onde nome, sobrenome e matrícula devem ser substituídos por suas informações pessoais.

5.3 Código

Você deve utilizar a linguagem C ou C++ para o desenvolvimento do seu sistema. O uso de estruturas pré-implementadas pelas bibliotecas-padrão da linguagem ou terceiros é terminantemente vetado. Você DEVE utilizar a estrutura de projeto abaixo junto ao Makefile:

```
– TP
  |– src
  |– bin
  |– obj
  |– include
  Makefile
```

A pasta **TP** é a raiz do projeto; **src** deve armazenar arquivos de código (`*.c`, `*.cpp`, ou `*.cc`); a pasta **include**, os cabeçalhos (headers) do projeto, com extensão `*.h`, por fim as pastas **bin** e **obj** devem estar vazias. O Makefile deve estar na raiz do projeto. A execução do Makefile deve gerar os códigos objeto `*.o` no diretório **obj** e o executável do TP no diretório **bin**. O arquivo executável **DEVE** se chamar **tp3.out** e deve estar localizado na pasta **bin**. O código será compilado com o comando:

```
make all
```

O seu código será avaliado através de uma **VPL** que será disponibilizada no moodle. Você também terá à disposição uma VPL de testes para verificar se a formatação da sua saída está de acordo com a requisitada. A VPL de testes não vale pontos e não conta como trabalho entregue. Um pdf com instruções de como enviar seu trabalho para que ele seja compilado corretamente estará disponível no Moodle.

¹Para este trabalho não é necessário analisar a localidade de referência.

6 Avaliação

- Corretude na execução dos casos de teste - (20% da nota total)
- Indentação, comentários do código fonte e uso de boas práticas - (10% da nota total)
- Conteúdo segundo modelo proposto na seção **Documentação**, com as seções detalhadas corretamente - (20% da nota total)
- Definição e implementação das estruturas de dados e funções - (15% da nota total)
- Apresentação da análise de complexidade das implementações - (5% da nota total)
- Análise experimental - (25% da nota total)
- Aderência completa às instruções de entrega - (5% da nota total)

Se o programa submetido **não compilar**² ou se compilar mas não passar em **pelo menos um caso de teste**, seu trabalho não será avaliado e sua nota será **0**. **Trabalhos não poderão ser entregues com atraso.**

7 Considerações finais

1. Comece a fazer esse trabalho prático o quanto antes, enquanto o prazo de entrega está tão distante quanto jamais estará.
2. Leia atentamente o documento de especificação, pois o descumprimento de quaisquer requisitos obrigatórios aqui descritos causará penalizações na nota final.
3. Certifique-se de garantir que seu arquivo foi submetido corretamente no sistema.
4. Plágio é crime. Trabalhos onde o plágio for identificado serão **automaticamente anulados** e as medidas administrativas cabíveis serão tomadas (em relação a todos os envolvidos). Discussões a respeito do trabalho entre colegas são permitidas. É permitido consultar fontes externas, desde que exclusivamente para fins didáticos e devidamente registradas na seção de bibliografia da documentação. **Cópia e compartilhamento de código não são permitidos.**

8 FAQ (*Frequently asked Questions*)

1. Posso utilizar qualquer versão do C++? NÃO, o corretor da VPL utiliza C++11.
2. Posso fazer o trabalho no Windows, Linux, ou MacOS? SIM, porém lembre-se que a correção é feita sob o sistema Linux, então certifique-se que seu trabalho está funcional em Linux.
3. Posso utilizar alguma estrutura de dados do C++ do tipo Queue, Stack, Vector, List, etc? NÃO.
4. Posso utilizar smart pointers? NÃO.

²Entende-se por compilar aquele programa que, independente de erros no Makefile ou relacionados a problemas na configuração do ambiente, funcione e atenda aos requisitos especificados neste documento em um ambiente Linux.

5. Posso utilizar o tipo String? SIM.
6. Posso utilizar o tipo String para simular minhas estruturas de dados? NÃO.
7. Posso utilizar alguma biblioteca para tratar exceções? SIM.
8. Posso utilizar alguma biblioteca para gerenciar memória? SIM.
9. As análises e apresentação dos resultados são importantes na documentação? SIM.
10. Os meus princípios de programação ligados a C++ e relacionados a engenharia de software serão avaliados? NÃO.
11. Posso fazer o trabalho em dupla ou em grupo? NÃO.
12. Posso trocar informações com os colegas sobre os fundamentos teóricos do trabalho? SIM.
13. Posso utilizar IDEs, Visual Studio, Code Blocks, Visual Code, Eclipse? SIM.