Universidade Federal de Viçosa – Campus UFV-Florestal
Ciência da Computação – Projeto e Análise de Algoritmos
Professor: Daniel Mendes Barbosa

Trabalho Prático 3

Este trabalho é **obrigatoriamente em grupo**. Os grupos já foram definidos nesta planilha e este trabalho deverá ser entregue no PVANet Moodle de acordo com as instruções presentes no final da especificação.

O reino de Hyrule está em perigo - de novo - e Link recebeu uma profecia de como restaurar o equilíbrio e salvar o dia. No entanto, o texto da profecia é ilegível! A princesa Zelda está igualmente confusa quanto ao significado, mas teve uma brilhante ideia: entrar em contato com outra linha do tempo, num futuro distante, no qual alguém será capaz de decifrar a mensagem. Ela acaba conseguindo falar com um grupo de estudantes de Ciência da Computação, e vocês a acalmam, pois trata-se de um esquema simples de criptografia - naquela época ainda não conheciam técnicas mais avançadas.

A tarefa de vocês então é realizar uma criptoanálise clássica para decifrar a profecia. O grupo deve criar um programa **interativo** na linguagem C que seja capaz de realizar algumas operações que fazem parte do processo de criptoanálise, e fornecer como saída final a chave de criptografia e o texto decifrado.

Ao longo de todas as seções a seguir, considere as seguintes informações como exemplos da chave de criptografia, do texto claro, e do texto criptografado.

Chave:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
S	T	G	B	H	V	L	K	Q	O	I	M	C	Z	R	E	A	X	Y	W	P	D	N	F	J	U

Texto claro:

```
O HEROI LINK PRECISA DA AJUDA DA PRINCESA ZELDA PARA QUEBRAR O CODIGO.
```

Texto criptografado:

```
R KHXRQ MQZI EXHGQYS BS SOPBS BS EXQZGHYS UHMBS ESXS APHTXSX R GRBQLR.
```

Como indicado pela chave, a letra O foi mapeada para a letra R, a letra H foi mapeada para a letra K, e assim por diante. **Cada grupo receberá um texto criptografado *diferente*, que de fato será usado no trabalho. Para receber este arquivo, um aluno de cada grupo deverá enviar um e-mail para o monitor da disciplina, no endereço henrique.s.santana@ufv.br, solicitando o envio do arquivo.** Além disso, o texto será consideravelmente mais longo que esse exemplo, para possibilitar, por exemplo, a análise de frequência.

Entrada

A entrada principal do programa é apenas um arquivo de texto contendo o texto criptografado. Considere que o texto claro era originalmente escrito em português, sem nenhuma acentuação. Caracteres especiais, como vírgulas, pontos finais, etc, foram deixados inalterados, sendo apenas as letras mapeadas, e estas se apresentam sempre em caixa alta. Além das palavras em português, pode haver alguns poucos nomes próprios.

Saída

A saída final do programa deve ser também em arquivos de texto, um para a chave de criptografia encontrada pelo grupo, e outro para o texto claro obtido ao final do processo de criptoanálise. Ao longo da execução do programa, outros resultados intermediários, a depender da operação realizada, também serão exibidos na tela (saída padrão) para o usuário, descritos nas seções seguintes.

Interface

Como citado anteriormente, o programa implementado deve ser interativo, e portanto precisa de uma interface mínima com o usuário. Primeiramente, o programa deve solicitar qual arquivo será lido para ser o texto criptografado, ou então o caminho do arquivo pode ser passado pela linha de comando (`argv`).

Em seguida, o programa deve entrar em loop, exibindo ao usuário todas as opções de operação a cada iteração. As operações são:

1. Apresentar o estado atual da criptoanálise;
2. Fazer análise de frequência no texto criptografado;
3. Realizar casamento exato de caracteres no texto criptografado;

4. Realizar casamento aproximado de caracteres no texto parcialmente decifrado;
5. Alterar chave de criptografia;
6. Exportar resultado e encerrar o programa.

Estado atual da criptoanálise

O programa deve manter, ao longo de toda sua execução, o texto criptografado e a chave de criptografia até então descoberta. Inicialmente, a chave estará vazia, pois não se sabe qual letra foi mapeada para qual:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Ao selecionar a opção 1, de exibir o estado atual da criptoanálise, o programa deveria então exibir: o texto criptografado, a chave - neste caso, vazia -, e novamente o texto, sem nenhuma alteração. Suponha que, após outras operações, houve a suspeita de que a letra A tenha sido mapeada para a letra S, e a letra E para H. Assim, agora a chave estará parcialmente preenchida, e o resultado desta operação 1 será diferente:

```
=== Texto criptografado ===
R KHXRQ MQZI EXHGQYS BS SOPBS BS EXQZGHYS UHMBS ESXS APHTXSX R GRBQLR.

=== Chave ===
ABCDEFGHIJKLMNOPQRSTUVWXYZ
S   H

=== Texto parcialmente decifrado ===
R KEXRQ MQZI EXEGQYA BA AOPBA BA EXQZGEYA UEMBA EAXA APETXAX R GRBQLR.
   ^           ^   ^  ^ ^   ^  ^      ^ ^  ^  ^  ^ ^   ^  ^
```

A versão parcialmente decifrada do texto também será exibida, usando algum recurso para marcar quais letras já foram trocadas, evitando confusão entre as partes criptografadas ou não. Recomenda-se o uso de algum caractere especial abaixo ou acima de cada letra trocada, ou, melhor ainda, uso de cores diferenciadas para as letras trocadas.

Análise de frequência

A opção 2 consiste em contar quantas vezes cada letra aparece no texto, e relatar o resultado da contagem ao usuário. A contagem deve ser exibida em ordem decrescente, podendo conter apenas a frequência absoluta das letras, ou também incluir a frequência relativa, como mostra o exemplo abaixo.

Por motivos de praticidade, o programa pode exibir também a tabela de frequência relativa das letras em português, seja ao lado ou abaixo da tabela do texto. É importante que ambas as tabelas sejam consultadas para que o grupo possa tentar adivinhar qual letra corresponde a outra. Por exemplo, sabe-se que a letra A é a mais frequente na língua

portuguesa, e na tabela de frequências das letras do texto, a letra S é a mais frequente. Seria possível então supor (corretamente) que A foi mapeada em S.

```
Letra,  Cont.,  Freq.
S          10  17.54%
X           6  10.53%
R           5   8.77%
H           5   8.77%
Q           5   8.77%
B           5   8.77%
E           3   5.26%
G           3   5.26%
Y           2   3.51%
Z           2   3.51%
M           2   3.51%
P           2   3.51%
I           1   1.75%
K           1   1.75%
O           1   1.75%
U           1   1.75%
A           1   1.75%
T           1   1.75%
L           1   1.75%
```

A formatação exata da tabela é de escolha do grupo. Independente se a tabela de frequências relativa da língua portuguesa for exibida ou não, o grupo deve colocá-la no relatório do trabalho, com uma referência de onde ela foi obtida.

Busca no texto criptografado

Esta operação consiste em realizar um casamento exato de caracteres sobre o texto criptografado. Após selecionar a opção 3, o usuário poderá digitar um padrão, e um algoritmo de casamento de caracteres será executado, contando quantas ocorrências do padrão existem no texto criptografado.

Por exemplo, suponha que vocês estão observando o texto criptografado e percebem que o par de letras “BS” aparece isolado e com frequência. Sabendo que a letra S era A antes da criptografia, vocês suspeitam que “BS” seja a palavra “DA”. Executando então a operação 3, usando “BS” como entrada, o programa informa a vocês:

```
Qual o padrão utilizado?
> BS
```

```
Ocorrências: 3
```

Essa é a saída mínima esperada para essa operação. Para tornar essa consulta mais útil, o grupo pode implementar como adicional: exibir onde ocorreu cada casamento do padrão, mostrar a frequência do padrão com relação a quantidade de palavras, mostrar a frequência do padrão com relação a quantidade de letras, ou outras métricas que o grupo achar relevante.

O objetivo desta operação é ajudar o grupo a decidir se algum par, trio, ou conjunto qualquer de letras, é frequente no texto criptografado, de forma a gerar suspeitas de como as letras foram mapeadas. Para implementá-la, o grupo pode decidir qual algoritmo de casamento de caracteres será usado: KMP, Shift-And, Boyer-Moore ou alguma variação destes. O próprio grupo deve implementar o algoritmo usado, e a escolha deve ser documentada no relatório.

Busca no texto parcialmente decifrado

Esta operação consiste em realizar um casamento aproximado de caracteres sobre o texto parcialmente decifrado. Ao selecionar a opção 4, o usuário poderá digitar um padrão, e um algoritmo de casamento de caracteres será executado, contando quantas ocorrências aproximadas do padrão foram encontradas.

O algoritmo usado nesta operação deve ser obrigatoriamente o Shift-And aproximado. Além do padrão, o usuário deve poder escolher qual a tolerância de erro no casamento aproximado ele deseja, ou seja, qual a distância de edição máxima entre o padrão e sua ocorrência no texto. Apenas o erro por substituição deve ser considerado, visto que o erro por inserção ou remoção não é relevante para este contexto.

Voltando ao exemplo, suponha que agora já se sabe também que a letra D foi mapeada em B, de forma que, ao executar a operação 1, o estado atual do programa é informado como segue:

```
=== Texto criptografado ===
R KHXRQ MQZI EXHGQYS BS SOPBS BS EXQZGHYS UHMBS ESXS APHTXSX R GRBQLR.

=== Chave ===
ABCDEFGHIJKLMNOPQRSTUVWXYZ
S  BH

=== Texto parcialmente decifrado ===
R KEXRQ MQZI EXEGQYA DA AOPDA DA EXQZGEYA UEMDA EAXA APETXAX R GRDQLR.
   ^           ^   ^ ^^ ^  ^^ ^^      ^ ^  ^ ^^  ^ ^   ^  ^      ^
```

Imaginando o contexto da mensagem, vocês poderiam supor que o nome da princesa seria mencionado. Das 5 letras que formam o nome dela, 3 já foram mapeadas, sobrando

2 letras ainda desconhecidas. Assim, a opção 4 é selecionada, o padrão usado é "ZELDA" e vocês decidem usar 2 como tolerância:

```
Qual o padrão e a tolerância utilizados?
> ZELDA 2

Ocorrências: 1
@[42,47): UEMDA
```

Como visto no exemplo, devem ser informadas, obrigatoriamente, a quantidade das ocorrências, além da posição e do texto contido em cada uma dessas ocorrências. O índice utilizado para as letras começa em zero, e o intervalo é fechado na posição inicial, e aberto na posição final. Opcionalmente, o grupo pode tornar essa saída mais visual, usando recursos como, por exemplo, a coloração dos caracteres.

O objetivo dessa operação é ajudar o grupo a procurar por palavras parcialmente decifradas no texto, de forma que seja possível inferir as letras que faltam numa palavra. No exemplo, poderia-se supor corretamente que a letra Z foi mapeada para U, e que a letra L foi mapeada para M.

Alterar a chave de criptografia

Após cada uma dessas operações de análise no texto, novas suposições podem ser feitas quanto à chave, e é através desta operação que o usuário vai realizar essas atualizações. Ao escolher a opção 5, o usuário poderá inserir duas informações: a letra original e a letra para o qual ela foi mapeada, e em seguida o programa relata a alteração feita.

Por exemplo, a primeira suposição feita foi de que a letra A foi mapeada para S. Nesse caso, essa seria a operação:

```
Informe a letra original, seguida da letra para a qual foi
mapeada:
> A S

Registrado: A -> S
```

Caso o grupo ache que seja mais conveniente especificar as letras na ordem contrária - primeiro a letra criptografada e depois a letra original -, essa decisão deve constar no relatório. O mesmo vale para a exibição da chave, na opção 1, caso o grupo ache melhor apresentá-la na ordem inversa.

Exportar resultado

Por fim, quando a criptoanálise estiver encerrada, a opção 6 deve solicitar ao usuário um caminho de arquivo para salvar a chave, e outro para salvar o texto decifrado, e logo em seguida encerrar o programa.

Faça exatamente o que está sendo pedido neste trabalho, ou seja, mesmo que o grupo tenha uma ideia mais interessante para o programa, deverá ser implementado exatamente o que está definido aqui no que diz respeito ao problema em si, às etapas da criptoanálise e os algoritmos especificados. O grupo pode implementar algo além disso, desde que não atrapalhe a obtenção dos resultados necessários a esta especificação.

Formato e data de entrega

Os arquivos com o código-fonte (projeto inteiro do Codeblocks ou arquivos .c, .h e makefile), juntamente com um arquivo PDF (**testado, para ver se não está corrompido**) contendo a **documentação**. A documentação deverá conter:

- explicação dos algoritmos projetados;
- implementação do algoritmo projetado (estruturas de dados criadas, etc);
- explicação de como compilar o programa.
- resumo das principais etapas realizadas pelo grupo para criptoanálise;
- exemplos de entrada e saída para cada operação contida nesse resumo;
- resultado final da chave encontrada e o texto decifrado.

Mais direcionamentos sobre o formato da documentação podem ser vistos no documento "Diretrizes para relatórios de documentação".

Importante: Entregar no formato **ZIP**. As datas de entrega estarão configuradas no PVANet Moodle. É necessário que apenas um aluno do grupo faça a entrega, mas o PDF da **documentação deve conter os nomes e números de matrícula de todos os alunos do grupo que efetivamente colaboraram com o trabalho em sua capa ou cabeçalho**.

Bom trabalho!