

CRIPTOANÁLISE

Virgilius Santos <virgilius.santos@acad.pucrs.br>

Pontifícia Universidade Católica do Rio Grande do Sul – Escola Politécnica – Curso de Engenharia de Software
Av. Ipiranga, 6681 Prédio 32 Sala 505 – Bairro Partenon – CEP 90619-900 – Porto Alegre – RS

23 de abril de 2020

RESUMO

Este artigo irá descrever como conseguimos decifrar um texto encontrando a chave usando o índice de coincidência e em seguida decodificando usando a cifra de Vigenère.

1 INTRODUÇÃO

O objetivo deste trabalho é aplicar os conhecimentos sobre criptoanálise.

Usando a linguagem Swift criamos um framework capaz de receber um texto codificado, buscar uma chave de criptografia usando o índice de coincidência e por fim decifrar o texto com essa chave.

2 ALGORITMO

O Algoritmo foi desenvolvido em torno de duas funções a primeira buscar a chave usada para codificar, e a segunda função usada para decodificar o texto.

2.1 *Buscar a Chave*

Essa função foi feita para, usando o índice de coincidências, encontrar a chave de encriptação.

Para fazer isso primeiramente ela precisa que o texto seja transformado em uma lista de caracteres A.

A seguir mostrarmos o algoritmo que usamos para contar a frequência dos caracteres:

```
para (começo: Inteiro = 0, começo < passo, começo ++){  
    dictArray = [[Char: Int]]  
    dict = [Char: Int]  
    dictArray.adicionar(dict)  
    para (posição: Inteiro = começo, posição < A.tamanho, posição += passo){  
        dict[A[posição]] += 1  
    }  
}
```

Esse dict é um dicionário onde a chave é um caractere encontrado no texto e o valor é o número de vezes que ele aparece.

O dictArray é uma lista formada com os dicionários gerados com a variação dos passos.

Um exemplo para a palavra “celularidade” e passo igual a 2:

A = [c, e, l, u, l, a, r, i, d, a, d, e], passo = 2;

dictArray = [[c: 1, l: 2, r: 1, d: 2], [e: 2, u: 1, a: 2, i: 1]]

Dessa forma garantimos que no dicionário de caracteres apareçam apenas os caracteres que ocorrem naquele passo usado.

Em seguida calculamos o índice de coincidência para cada dicionário da lista de dicionários e tiramos uma média para saber o valor do índice de coincidência daquele passo.

Para fazer o cálculo do índice de coincidência usamos a função a seguir:

$$IC = \frac{\sum_{i=1}^c n_i(n_i - 1)}{N(N - 1)/c}$$

Figura 1 - índice de coincidência

Onde n_i é cada valor do dict, N é o somatório de todos os valores do dict e $c = 26$ que é o número de caracteres do nosso alfabeto. Para facilitar usamos sem normalizar o denominador, ou seja, não dividimos o denominador por c (Índice da coincidência, Wikipédia 2020).

Para encontrarmos a chave fazemos o cálculo dos índices de coincidência para passos variando de 1 a 10 e escolhemos o que mais se aproxima do valor ideal, para textos em inglês o valor é 0.0667 e para textos em português o valor é 0.072723.

Ao determinarmos o índice de coincidência mais próximo, pegamos da lista de dicionários de caracteres daquele índice o caracter mais frequente e convertemos ele usando o caractere mais frequente da língua, no caso do inglês e português usamos “e” junto com a matriz de Vigenère inversa, explicaremos essa matriz a seguir.

2.2 Matriz Vigenère Inversa

Para facilitar o trabalho dentro de algoritmo na hora de decodificar o texto, geramos uma matriz de Vigenère onde os valores estão como índices das colunas e os índices estão como valores na matriz.

A matriz normal seria como a figura mostrada a seguir:

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

Figura 2 - Matriz Vigenère

Já na matriz invertida seria como a imagem a seguir:

	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
A	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
B	z	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
C	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
D	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
E	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
F	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
G	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
H	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
I	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
J	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
K	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
L	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
M	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n
N	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m
O	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l
P	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k
Q	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j
R	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i
S	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
T	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
U	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
V	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e
W	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d
X	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c
Y	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b
Z	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	a

Figura 3 - Matriz Vigenère Invertida

Essa matriz permite que se eu tiver uma letra da chave e o valor cifrado consigo decifrar mais rapidamente.

2.3 Decodificação

A decodificação do texto é feita parecida com a decodificação da chave, mas no lugar do caractere mais frequente, usamos a chave, ou seja, para cada valor do texto, usamos a chave daquela posição para decodificar.

3 CONCLUSÃO

Neste artigo usamos o índice de coincidência para encontrar o tamanho da chave e a letra mais frequente da língua do texto para decodificar a chave, usando esta chave decodificamos todo o texto e imprimimos o terminal.

Durante o desenvolvimento vimos que não era uma boa prática partir o texto em trechos menores, porque isso impactava diretamente a performance do algoritmo, já que era criadas várias palavras em memória. Pra resolver esse problema, convertemos o texto pra uma lista de caracteres e apenas percorremos esta lista.

Outro ponto importante é que para ver as letras mais frequentes estava criando dicionários com todas as letras do alfabeto, mas ao fazer isso também pesava durante o processo para gerar os índices de coincidência.

Por fim o ultimo ganho veio com a matriz de Vigenère invertida, que reduziu a decodificação para uma busca $O(1)$.

Para confirmar os resultados usamos algumas ferramentas online e também criamos testes unitários tanto para validarem as funções criadas, quando para verificar a performance do algoritmo, sendo que para um texto grande disponibilizado o tempo caiu de alguns minutos partindo o texto em pedaços menores, para menos de 4 segundos quando passamos a percorrer a lista de caracteres.

4 REFERÊNCIA:

https://pt.qwe.wiki/wiki/Index_of_coincidence