

# Pregunta G Interrogación IIC2513 2021-1

## SECCIÓN: Programación 2

### Descifrado de encriptación por sustitución

La encriptación por sustitución es un caso básico de encriptación muy usado en la antigüedad pero que ya es vulnerable, al punto que no se usa (en forma seria).

El mecanismo es sencillo: a cada letra del abecedario (es decir: a,b,c,d,...w,x,y,z) se le asigna arbitrariamente otra letra del abecedario. El único cuidado a considerar es no repetir letras en la asignación. Así pues, se podría realizar, por ejemplo, la siguiente tabla de traducción:

a	b	c	d	e	f	.....	w	x	y	z
t	m	ñ	u	c	a		z	b	l	q

Para cada columna encontramos: en la primera fila, la letra del abecedario como lo conocemos (incluimos ñ y w). En la segunda fila, la letra que se usará para reemplazar la de la primera fila en nuestro mensaje encriptado.

Según la tabla resumida que se expone como ejemplo, la palabra **cada** se encriptará como: **ñtut**

Existe una forma de descifrar lo anterior, basada en el hecho de que las letras son usadas con distinta frecuencia en nuestro idioma. Se puede construir una tabla de frecuencias de cada letra en el abecedario usual. Luego, se puede hacer lo mismo para las letras del mensaje encriptado y comparar para “mapear” las equivalencias y construir la tabla de traducción.

Esta es la tabla de frecuencia de letras para el idioma español:

Letra	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Porcentaje	12,53%	1,42%	4,68%	5,86%	13,68%	0,69%	1,01%	0,70%	6,25%	0,44%	0,02%	4,97%	3,15%	6,71%
Letra	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	
Porcentaje	0,31%	8,68%	2,51%	0,88%	6,87%	7,98%	4,63%	3,93%	0,90%	0,01%	0,22%	0,90%	0,52%	

En el caso particular de un texto de longitud apreciable, como el Quijote, la frecuencia de aparición de las letras es:

Letra	e	a	o	s	n	r	i	l	d	u	t	c	m	p
Cantidad	229.188	200.492	162.512	125.726	108.440	100.953	90.070	89.141	87.237	79.471	61.749	59.435	44.658	35.464
Porcentaje	14,0%	12,2%	9,9%	7,7%	6,6%	6,2%	5,5%	5,4%	5,3%	4,8%	3,8%	3,6%	2,7%	2,2%
Letra	q	y	b	h	v	g	j	f	z	ñ	k	w	x	
Cantidad	32.483	25.115	24.146	19.920	17.855	17.225	10.530	7.581	6.491	4.241	377	2	1	
Porcentaje	2,0%	1,5%	1,5%	1,2%	1,1%	1,0%	0,6%	0,5%	0,4%	0,3%	0,1%	0,0%	0,0%	

(Fuente: Wikipedia)

Como se puede apreciar, el caso de la letra “e” representa el 14% de las letras en el Quijote, y en la tabla de frecuencias, la “e” es el 13,68%. En el caso de la letra “ñ”, aparece 0,3% en el Quijote, mientras que en la tabla de frecuencia es el 0,31%.

### SE PIDE:

Construir `decryptText(text, freq)` donde la función `decryptText()` recibe dos argumentos: `text`: un texto encriptado por sustitución y `freq`; un objeto con la frecuencia de letras. La función debe retornar un texto (string) con el texto descifrado.

Para el propósito de la pregunta, ***freq*** es el siguiente arreglo de objetos (ÚSELO **obligatoriamente en su respuesta**)

```
freq = [ {l:"a", f:12.53}, {l:"b", f:1.42}, {l:"c", f:4.3}, {l:"d", f:4.68}, {l:"e", f:13.68}, {l:"f", f:1.02}, {l:"g", f:1.07}, {l:"h", f:1.0}, {l:"i", f:4.85}, {l:"j", f:0.44}, {l:"k", f:0.02}, {l:"l", f:5.46}, {l:"m", f:2.75}, {l:"n", f:7.26}, {l:"ñ", f:0.31}, {l:"o", f:8.68}, {l:"p", f:2.51}, {l:"q", f:1.03}, {l:"r", f:6.87}, {l:"s", f:6.68}, {l:"t", f:4.63}, {l:"u", f:3.93}, {l:"v", f:0.77}, {l:"w", f:0.01}, {l:"x", f:0.22}, {l:"y", f:1.01}, {l:"z", f:0.52} ];
```

Donde ***l*** representa la letra y ***f*** la frecuencia de la letra.

### Consideraciones:

1. Puede utilizar todos los métodos de **Math**, **Array** y **String** que considere necesarios.

2. Puede crear otras funciones y estructuras de datos que considere necesarias.
3. Modificamos las frecuencias para que pudiesen probar el ejemplo de control que les entregamos (para que puedan ver que su desarrollo funciona adecuadamente). Debe saber que en general (y en particular en el ejemplo) **no hay una coincidencia exacta de frecuencias** que encontrará en el texto versus las frecuencias entregadas anteriormente, por lo tanto su programa deberá encontrar la frecuencia que más se adecua sin ser exacta (que es lo más común)
4. De lo anterior se desprende que la traducción puede no ser exacta, es decir, puede haber letras intercambiadas, sin embargo, debe poder leerse o interpretarse el texto. Sobre todo, las letras más frecuentes deben poder ser identificadas sin problemas.
5. Dado que la frecuencia no distingue mayúsculas de minúsculas, usted deberá transformar el texto a minúsculas
6. No hay letras acentuadas. **Además, considere que solo se dejaron en el texto los siguientes símbolos de puntuación: “,” coma, “;” punto y coma y “.” punto.** Todos los otros símbolos no se usarán (ni signos de interrogación, ni paréntesis, ni otros). Debe considerar que sí existen los espacios en blanco.
7. Puede que, al momento de decodificar, tenga varias letras asociadas al texto que son “candidatos” para la decodificación. Debe tener cuidado al respecto, pues no puede repetir letras en su decodificación. Así pues, si determina que una letra X es la adecuada para decodificar una cierta letra Y, esa misma letra X no puede ser usada para decodificar otra letra, por ejemplo Z (*con Z distinto a Y*), por lo que tendrá que usar otra letra para decodificar Z.
8. Su función se probará usando la estructura **freq** y el texto encriptado que se incluye más adelante en el enunciado.
9. *Insistimos que el texto debe quedar legible, no necesariamente exacto. Por ejemplo, una frase como 'sialrt ntvtnlr xp naimpm' podría ser traducida por su programa como “juiero conocer ya ciudad” y se consideraría un buen descifrado del mensaje ya que un humano es capaz de leerlo.*

El texto (presentado acá en la variable **text**) le servirá para probar su función. Este texto corresponde al fragmento inicial de un libro (¿Puede identificar el libro del que se extrajo el ejemplo?) que fue codificado utilizando el algoritmo de sustitución.

```
text = "woymohjñjsñ ñ pñ nñuñ hñohñi fwywi nw tñzsñ
zñihñxm ñimnñjnw, fsoswoxm lmj pñ yñppw xw ñp ñjym kgw
xñ ñ pñ lpñbñ xw ñjnñi c ñlwoñi pñ pgb xw ywosbñ cmpsfm
kgw vpmhñ imzjw wp jsm nw xwrñzñ xsihsougsj pñi vmjnñi,
cñ ig ispgwhñ xwpuñxñ iw soiyjszsñ wo wp lgwohw xw pñi
ñjhwi, ñ fwywi ñoxñoxm xw go pñxm ñ mhjm, ñ fwywi
xwhwosxñ wo wp ljwhsp xw tswjjm, soyppsoñxñ imzjw wp
ñugñ. c wjñ hño oñhgjñp yjgbñj pñ yñppw, igzsj pmi
lwpñami xwp lgwohw, wohjñj wo ig xwpuñxñ ysohgjñ c
ñywjyñjnw ñ pñ nñuñ kgw imojwsñ iso imjljwiñ,
```

ymofwoysxñ ynm cm xw kgw go woygwohjm yñigñp wñ pm  
nwomi yñigñp wo ogwihjñi fsxñi, c kgw pñ uwohw kgw iw  
xñ yshñi ljwysiñi wi pñ nsinñ kgw otywishñ lñlwp jñcñxm  
lñjñ wiyjszsjiw m kgw ñljswhñ xwixw ñzñrm wp hgzm xw  
xwohsvjsym. lwjm wppñ om wihñjsñ ñtmjñ wo wp lgwohw. ig  
vsoñ yñjñ xw hjñoipgysxñ lswp iw ñimnñjsñ ñ fswrmi  
lmjhñpwi wo wp utwhhm xwp nñjñsi, kgsbñ wihgfswjñ  
ytñjpñoxm ymo goñ fwoxwxmjñ xw lñlñi vjshñi m ymnswoxm  
goñ iñpytsytñ yñpswohw wo wp zmgpwfñjx xw iwzñihmlmp.  
xw hmxñi nñowjñi igzs tñihñ wp lgwohw, c om wihñzñ  
ñtmjñ om wihñzñ wo ns yñnsom, c ñgokgw ymomyíñnmi  
ogwihjmi xmnsyspsmi, yñxñ tgwym xw ogwihjñi xmi  
tñzshñysmowi xw vñpimi wihgxsñohwi wo lñjsi, yñxñ  
hñjrwhñ lmihñp ñzjswoxm goñ fwohñoshñ ymohjñ pñi  
nmpxgjñi zñjñhñi c pmi lñlwpwi ytsppmowi, ñgo ñis om  
omi zgiyñjsñnmi wo ogwihjñi yñiñi. ljwvwjsñnmi  
woymohjñjomí wo wp lgwohw, wo pñ hwjjñbñ xw go yñvw, wo  
go ysow ypgz m ñuñytñxmi rgohm ñ go uñhm wo ygñpkgswj  
lñhsm xwp zñjjsm pñhsom. ñoxñzñnmi iso zgiyñjomí lwjm  
iñzswoxm kgw ñoxñzñnmi lñjñ woymohjñjomí. wo yñxñ ngrwj  
lñjwysxñ ñ fmi iw ñumplñzñ ynm go ispwoysm  
woimjxwywxmj, goñ lñgiñ vspmiñ c yjsihñpsoñ kgw ñyñzñzñ  
lmj xwjggnzñjiw hjsihwnwohw, ynm go lñjñugñi nmrñxm  
kgw iw yswjjñ. rgihñnwohw go lñjñugñi, nñuñ, hw  
ñymjxñjsñi kgsbñ xw ñkgwp lñjñugñi fswrm kgw  
iñyjsvsysñnmi wo go zñjjñoym xwp lñjkgw, go ñhñjxwywj  
twpñxm xw nñjbm. pm hsjñnmi lmjkgw pm tñzsñi woymohjñxm  
wo pñ lpñbñ xw pñ ymoymjxsñ, cñ go lmym jmhm, c pm  
giñihw ngytsism, imzjw hmxm lñjñ nwhwjpm wo pñi  
ymihspñi xw pñ uwohw wo wp nwhjm c wo pmi ñghmzgiwi,  
iswnljw hmjlw c xsihjñsxñ c lwoiñoxm wo lñrñjmi lsohmi  
m wo go xszgrshm kgw tñysño xmi nmiyñi wo wp hwytm xwp  
ymytw, c ñkgwppñ hñjxw yñcm go ytñlñjjmo c kgsisihw  
ñzjsj mjugppmiñ hg lñjñugñi ygñoxm wohjñzñnmi wo

wplñjkgw, c wo hg nñom iw ñjnm goñ yñhñihjmvw xw  
jwpñnlñumi vjsmi c ogzwi owujñi, rsjmowi xw hwpñ  
xwihjmbñxñ yñcwoxm wohjw xwihwppmi xw fñjspñi  
xwiwoyñrñxñi, c omi jwsñnmi ynm pmym nswohjñi omi  
wnlñlñzñnmi, lwoiñoxm kgw go lñjñugñi woymohjñxm wo goñ  
lpñbñ xwzñi nmjsj xsuoñnwohw wo go lñjkgw, om lmxsñ  
wohjñj wo wp ysypm soomzpw xwp hñytm xw zñigjñ m xwp  
ymjxmo xw pñ fwjwxñ; wohmoywi cm pm ñjjmppw pm nwrmj  
lmiszpw, pm ppwfñnmi tñihñ pm ñphm xwp lñjkgw, ywjyñ  
xwp lgwohwysm imzjw wp vwjjmyñjjsp, c xwixw ñpps pm  
hsjw ymo hmxñi nsi vgwjbñi ñp vmoxm xw pñ zñjjñoyñ xw  
ywilwx nmrñxm nswohjñi ljmvwjsñi go ujshm xmoxw  
fñuñnwohw yjws jwymomywj goñ snljwyñysmo xw eñpqcjsñ.  
wo vso, om wi vñysp tñzpñj xw pñ nñuñ kgw ñ wihñ tmjñ  
ñoxñ iwugjñnwohw lmj nsjñoxm ñlpsyñxñnwohw wp igwpm  
tñihñ woymohjñj go lwxñbm xw uwowjm jmrn. is om pm  
woygowhjñ iwugsjñ ñis hmxñ pñ omytw, jwfmpfwjñ wo  
pmihñytm xw zñigjñ, pmi mrm fsexsmimi, ymofwoysxñ xw  
kgw ñpum tmjjszpw pw fñ ñ mygjjsj is om woygowhjñ wiñ  
ljwoxñ xw jwiynñhw, pñ iwañp xwp lwjxmo m xwp  
ñlpñbñnswohm. iw pm kgw wi wim lmjkgw hñnzso mzxwbyw  
ñ wiñi iwañpwi, hñnzso tñc fwywi wo kgw nw hmyñ  
woymohjñj hjñlm jmrn. xwixw pñ sovñoyñ ñlwoñi iw nw  
yñw ñpum ñp igwpm hwoum kgw pwfñohñjpm, iwñ pm kgw iwñ,  
lmjkgw is om pm tñum fñ ñ mygjjsj goñ xwiujñysñ, om ñ  
ns isom ñ ñpugswo ñ kgswo ñnm c ygcm omnzjw wnlswbñ ymo  
pñ sosysñp xwp mzwrm yñsxm. pm lwmj wi kgw oñxñ lgwxw  
ymohwowjnw ygñoxm ñpum iw nw yñw ñp igwpm, os hñnlmym  
fñpw kgw pm pwfñohw mhjm lmjkgw wp ññpwvsysm mzzñjsñ  
sugñp. tw lñiñxm ngytñi fwywi lmj pmym ñ yñgiñ xw wihm  
c pñ fwjxñx wi kgw wihmc pmym ygñoxm pm tñum, ygñoxm nw  
ljwyslsh ñ rgohñj go pñlsb m go hjmyshw xw lñlwp kgw  
iw nw tño sxm xw pñ nñom, ynm pñ omytw xwp hwjjmo xw  
ñbgyñj wo wp jwihñgjñohw xw pñ yñppw iyjszw, go

jwihñgjjñohw zñyño ymo nmohmowi xw uwjwohwi, lghñi xw  
bmjjmi lpñhwñxmi c nñhjsnmosmi zsw mjuñosbñxmi.  
wihñzñnmi ymo c ñ ns iw nw yñcm go hwjjmo xw ñbgyñj kgw  
vgw ñ lñjñj ñzñrm xw goñ nwiñ zñihñohw pwrmi xw pñ  
ogwihjñ. pm ljsnwjm kgw nw ppñnm pñ ñhwoysmo vgw pñ  
vmjnñ wo kgw wp hwjjmo iw tñzsñ ñpwrñxm, lmjkgw wo  
uwowjñp pmi hwjjmowi xw ñbgyñj iw lpñohño ñlwoñi hmyño  
wp igwpm lmj jñbmowi lñjñpwpwlslwñi wfsxwohwi. lwjm  
wihw iw ymoxgysñ ymn is vgwjñ goñ zmpñ xw oñvhñpsoñ,  
pm ygñp ñgnwohm ns ñljwoismo, c ppwugw ñ yjwwj kgw  
jwñpnwohw nw pm tñzsño ñjjñoyñxm xw pñ nñom. jmoñpx,  
kgw nw ymomyw, nsjm tñysñ xmoxw tñziñ sxm ñ lñjñj wp  
hwjjmo c iw wnlwbm ñ jwíj. wim nw xsm hmxñfíñ nñi  
nswxm, nwbypñxm ymo jñzsñ. go nmbm iw ñywjym lwoiñoxm  
kgw iw nw tñzsñ yñsxm ñpum ljwysmim, goñ lñjqwj m goñ  
xwohñxgjñ lmihsbñ, c wo jwñpsxñx pm gosym kgw tñysñ wjñ  
nmpwihñjnw, wohmoywi iso lwxsj lwjnsim nw hsjw ñp igwpm  
c wnlwyw ñ zgiyñj wp hwjjmo wohjw pmi bñlñhmi xw pñ  
uwohw kgw wihñzñ ppwoñ xw ygjsmisxñx yjwcwoxm c ymo  
jñbmo kgw iw hjñhñzñ xw ñpum snlmjhñohw"