educación IT
CODE YOUR FUTURE

Criptografía y Blockchain

Módulo 1

Criptoanálisis

Estadísticas del lenguaje

Lo primero que debe intentar un criptoanalista es un ataque a partir de las **estadísticas del lenguaje.**

El análisis de las **frecuencias relativas de aparición de caracteres** en el criptograma podría indicarnos el idioma original del texto claro.

Ejemplo

La letra **E** es la más frecuente en inglés, francés, alemán o castellano.

La diferencia porcentual con la segunda letra nos permitiría distinguir los idiomas.

- En inglés, la segunda letra (**t**) aparece un 29% menos.
- En francés, la segunda letra (**s**) aparece un 46% menos.
- En alemán, la segunda letra (**n**) aparece un 44% menos.
- En castellano, la segunda letra (**a**) aparece un 8%.

Repitiendo el análisis para más letras, podríamos distinguir el francés del alemán.

	Inglés	Francés	Alemán	Castellano
A	8.167%	7.636%	6.51%	12.53%
B	1.492%	0.901%	1.89%	1.42%
C	2.782%	3.260%	3.06%	4.68%
D	4.253%	3.669%	5.08%	5.86%
E	12.702%	14.715%	17.40%	13.68%
F	2.228%	1.066%	1.66%	0.69%
G	2.015%	0.866%	3.01%	1.01%
H	6.094%	0.737%	4.76%	0.70%
I	6.966%	7.529%	7.55%	6.25%
J	0.153%	0.545%	0.27%	0.44%
K	0.772%	0.049%	1.21%	0.00%
L	4.025%	5.456%	3.44%	4.97%
M	2.406%	2.968%	2.53%	3.15%
N	6.749%	7.095%	9.78%	6.71%
O	7.507%	5.378%	2.51%	8.68%
P	1.929%	3.021%	0.79%	2.51%
Q	0.095%	1.362%	0.02%	0.88%
R	5.987%	6.553%	7.00%	6.87%
S	6.327%	7.948%	7.27%	7.98%
T	9.056%	7.244%	6.15%	4.63%
U	2.758%	6.311%	4.35%	3.93%
V	0.978%	1.628%	0.67%	0.90%
W	2.360%	0.114%	1.89%	0.02%
X	0.150%	0.387%	0.03%	0.22%
Y	1.974%	0.308%	0.04%	0.90%
Z	0.074%	0.136%	1.13%	0.52%

En Inglés

- Conjuntos de dos letras: TH, HE, AN, IN, ER, RE, ES, ON, EA, TI, AT, ST, EN, ND, OR, TO, NT, ED, IS, AR.
- Conjuntos de tres letras: THE, ING, AND, ION, ENT, FOR, TIO, ERE, HER, ATE, VER, TER, THA, ATI, HAT, ERS.

En Francés

- Conjuntos de dos letras: ES, EN, OU, DE, NT, TE, ON, SE, AI, IT, LE, ET, ME, ER, EM, OI, UN, QU.
- Conjuntos de tres letras: ENT, QUE, ION, LES, AIT, TIO, ANS, ONT, ANT, OUR, AIS, OUS.

En Alemán

- Conjuntos de dos letras: EN, ER, CH, DE, GE, EI, IE, IN, NE, ND, BE, EL, TE, UN, ST, DI, NO, UE, SE, AU, RE, HE.
- Conjuntos de tres letras: EIN, ICH, DEN, DER, TEN, CHT, SCH, CHE, DIE, UNG, GEN, UND, NEN, DES, BEN, RCH.

En Castellano

- Conjuntos de dos letras: ES, EN, EL, DE, LA, OS, AR, UE, RA, RE, ER, AS, ON, ST, AD, AL, OR, TA, CO.
- Conjuntos de tres letras: QUE, EST, ARA, ADO, AQU, DEL, CIO, NTE, OSA, EDE, PER, IST, NEI, RES, SDE.

Método de Kasiski - Babbage

- **Sirve para atacar el cifrado de Vigenère.**
- Fue descubierto primero por Charles Babbage en 1858. Nunca lo hizo público.
- Friedrich Kasiski, sin conocer el trabajo de Babbage, lo redescubrió en 1863.
- El método de Vigenére tenía su problema con las repeticiones. Si la longitud del texto era suficientemente larga, podía comenzar a descubrirse a través de palabras comunes o repetidas.
- Si bien cada letra se sustituye por una diferente en cada aparición, lo cierto es que a través de esas palabras repetidas se puede ir descubriendo la clave, comenzando por su longitud.
- La probabilidad de que sean las mismas letras con las mismas partes de la clave aumenta a medida que ese texto repetido es más largo.
- A mayor longitud de la secuencia repetida en el texto cifrado, mayor es la probabilidad de repetición también.

Ejemplo de uso

Supongamos que tenemos un mensaje cifrado con Vigenére.

1. Primero lo separamos en bloques de 5 caracteres y buscamos los n-gramas repetidos:

```
PBVRQ VICAD SKAÑS DETSJ PSIED BGGMP SLRPW RÑPWY EDSDE ÑDRDP
CRCPQ MNPWK UBZVS FNVRD MTIPW UEQVV CBOVN UEDIF QLONM WNUVR
SEIKA ZYEAC EYEDS ETFPH LBHGU ÑESOM EHLBX VAEEP UÑELI SEVEF
WHUNM CLPQP MBRRN MPVIÑ MTIBV VEÑID ANSJA MTJOK MDODS ELPWI
UFOZM QMVNF OHASE SRJWR SFQCO TWVMB JGRPW VSUEX INQRS JEUEM
GGRBD GNNIL AGSJI DSVSU EEINT GRUEE TFGGM PORDF OGTSS TOSEQ
OÑTGR RYVLP WJIFW XOTGG RPQRR JSKET XRNBL ZETGG NEMUO TXJAT
ORVJH RSFHV NUEJI BCHAS EHEUE UOTIE FFGYA TGGMP IKTBW UEÑEN
IEEU
```

2. Luego, medimos la distancia de separación entre los tetragramas.

Obtenemos

- 3 cadenas GGMP separadas por 256 y 104 caracteres.
- 2 cadenas YEDS separadas por 72 caracteres.
- 2 cadenas HASE separadas por 156 caracteres.
- 2 cadenas VSUE separadas por 32 caracteres.

3. Estimamos la longitud probable de la clave como el máximo común divisor de las distancias calculadas previamente:

   L = mcd (32, 72, 104, 156, 256) = 4

4. Separemos el criptograma letras con ese salto:

   - Primer criptograma: letras 1, 5, 9, etc.
   - Segundo criptograma: letras 2, 5, 19, etc.

   Repetimos lo mismo para los otros dos.

C_A= PQAAEPDMRÑEEDCNUSRIECNIONSAAETLUOLAUIEULMNIIEAAOOLU
MNARSOMRSISERNAISIRTMDTOORLIORRENENOAVSNIAEOFAMTEO

C_B= BVDÑTSBPPPDÑPPPBFDPQBUFNUEZCDFBÑMBEÑSFNPBBÑBÑNMMKDPF
QFSJFTBPUNJMBNGDUUFPFSSÑRPFTPJTBTETTJFUBSUTFTPBÑE

C_C= VISSSIGSWWSDCQWZNMWVOEQMVIYESPHEEXEEEWMQRPMVISTMSWO
MOEWQWJWEQEGDISSETEGOOSETYWWGQSXLGMXOHHECEEIGGIWEE

C_D= RCKDJEGLRYDRRMKVVTUVVDLWRKEYEHGSHVPLVHCPRVTVDJJDEIZ
VHSRCGVXRUGGLJVEGEGRGTQGVJXGRKRZGUJRRVJHHUEEYGKUNU

5. Escribimos la cantidad de veces que aparece cada caracter en cada criptograma. Teniendo en cuenta que las letras más frecuentes del idioma español son la A,E,O,S, sabemos que la distancia entre ellas es (4,11, 4, 7).

Por simple inspección vemos que ese patrón de distancias se repite en cada criptograma, por lo que deducimos que el idioma del texto claro es castellano.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
C_A	11	0	2	3	12	1	0	0	11	0	0	5	6	9	1	10	2	1	9	7	4	5	1	0	0	0	0
C_B	0	14	1	6	4	12	0	0	0	4	1	0	3	6	8	6	14	2	1	6	9	7	1	0	0	0	1
C_C	0	0	1	2	18	0	7	3	7	1	0	1	7	1	0	0	2	6	1	12	3	0	3	12	3	2	1
C_D	0	0	3	5	7	0	12	6	1	7	5	4	1	1	0	6	2	1	13	2	3	7	14	0	2	3	2

6. Ahora nos paramos en las letras remarcadas en el punto anterior, (por ejemplo AEOS del primer criptograma) hacemos la suma de las repeticiones con el salto (4, 11, 4, 7).

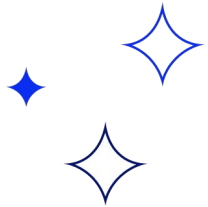
Repitiendo para todas las letras, los valores máximos en cada criptograma nos arrojan una letra de la clave K=ABER (no se muestran todos los valores).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
C1	11	0	2	3	12	1	0	0	11	0	0	5	6	9	1	10	2	1	9	7	4	5	1	0	0	0	0
C2	0	14	1	6	4	12	1	0	0	4	1	0	3	6	8	6	14	2	1	6	9	7	1	0	0	0	1
C3	0	0	1	2	18	0	7	3	7	1	0	1	7	1	0	0	2	6	1	12	3	0	3	12	3	2	1
C4	0	0	3	5	7	0	12	6	1	7	5	4	1	1	0	6	2	1	13	2	3	7	14	0	2	3	2
C1	40															20											
C2		49				25											27										
C3					51															28				13			
C4							27												45				21				

Resultado:

Si aplicamos Vigenère, obtenemos el texto descifrado:

"*Para que la cosa no me sorprenda como otros años he comenzado ya con unos suaves ejercicios de precalentamiento, mientras desayunaba he contemplado una bola plateada y una tira de espumillón y mañana me iniciaré en el amor al prójimo con los que limpien el parabrisas en los semáforos. Esta gimnasia del corazón metafórico es tan importante como la del otro corazón porque los riesgos coronarios están ahí escondidos tras la vida sedentaria y parapetados en fechas como estas de Navidad*".



La máquina Bomba de Turing-Welchman

Fue diseñada para **descubrir algunos de los ajustes diarios de las máquinas Enigma** en las diversas redes militares alemanas: específicamente, el conjunto de rotores en uso y sus posiciones en la máquina; las posiciones de inicio de los rotores y los cableados del *plugboard.*

Era un dispositivo electromecánico que replicaba la acción de varias máquinas Enigma conectadas entre sí. **La bomba británica estándar equivalía a 36 máquinas Enigma.**

Durante la Segunda Guerra Mundial se construyeron más de 200 bombas, pero fueron completamente destruidas al final de la misma.

Herramientas modernas

dCode

- El sitio web dCode decodifica automáticamente una amplia variedad de cifrados.
- Posee un detector de códigos que **reconoce automáticamente más de 200 tipos de codificación.**
- dCode es una enorme biblioteca de *scripts* para decodificar o codificar mensajes con técnicas de criptografía clásicas y modernas.

CyberChef

- La herramienta CyberChef está disponible en el sitio web.
- También podemos descargarla desde ese mismo enlace y usarla en nuestro equipo.
- Contiene cientos de operaciones posibles para **análisis y manipulación de datos, criptografía, compresión, *hashing*,** etc.
- Es desarrollada y mantenida por el GCHQ, uno de los tres servicios de inteligencia del Reino Unido.

¡Sigamos trabajando!