Criptografía y criptoanálisis clásicos

[2.1] ¿Cómo estudiar este tema?

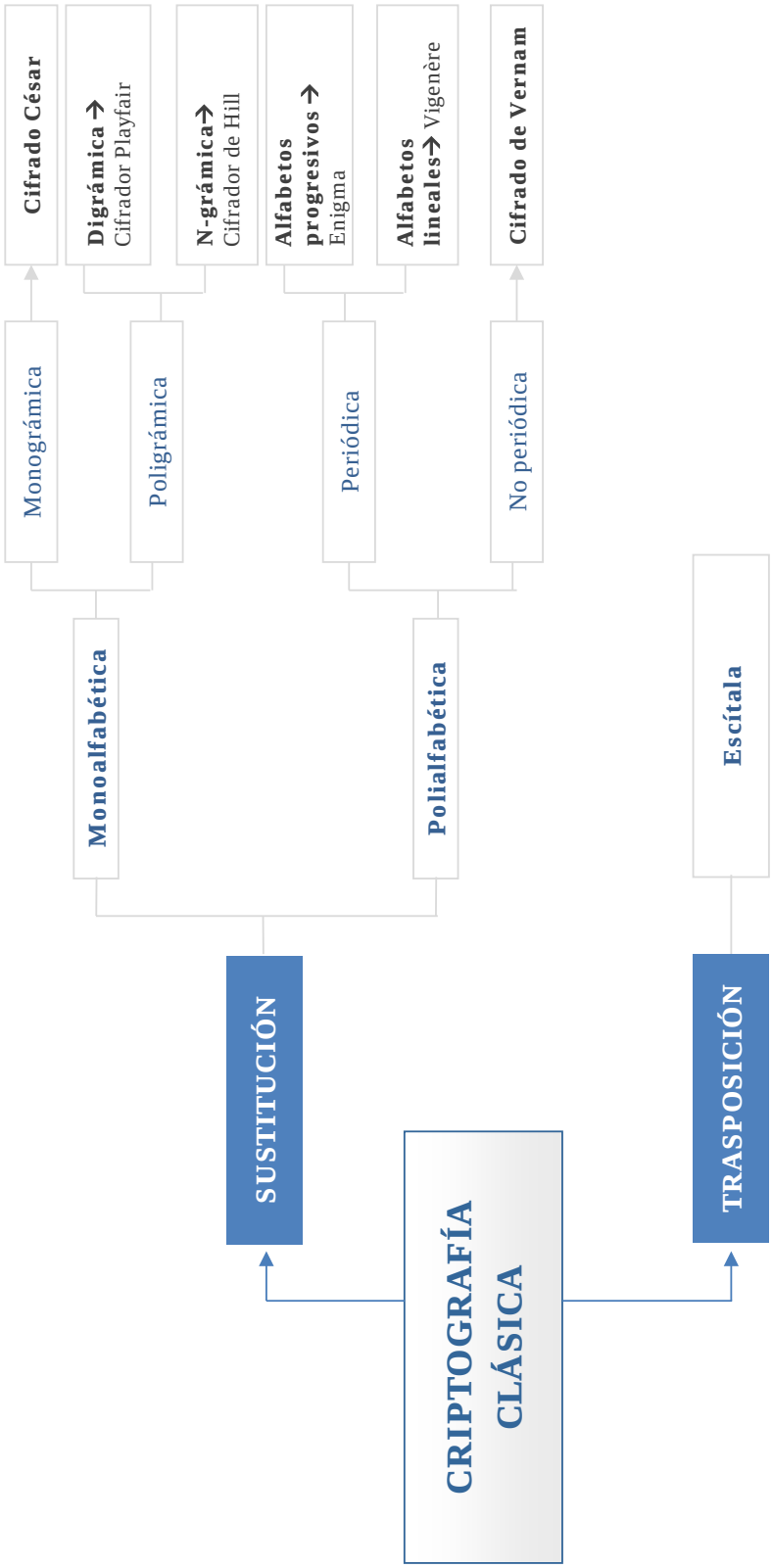
[2.2] Historia de la criptografía

[2.3] Cifradores de sustitución

[2.4] Caso de estudio: la *máquina Enigma*

TEMA 2

Esquema


CRIPTOGRAFÍA CLÁSICA
SUSTITUCIÓN
TRASPOSICIÓN
Monoalfabética
Polialfabética
Escítala
Monográmica
Polígrámica
Periódica
No periódica
Cifrado César
Digrámica → Cifrador Playfair
N-grámica→ Cifrador de Hill
Alfabetos progresivos → Enigma
Alfabetos lineales→ Vigenère
Cifrado de Vernam

Ideas clave

2.1. ¿Cómo estudiar este tema?

El estudio de este tema se realiza a través de los contenidos desarrollados en las **Ideas clave** expuestas a continuación.

En este tema estudiaremos los **métodos clásicos de cifrado**, que son aquellos utilizados desde prácticamente la invención de la escritura hasta la aparición de los ordenadores y los microprocesadores. Estos métodos se distinguen principalmente por usar simples **técnicas de sustitución y transposición de los caracteres del texto «en claro»**. Todos los métodos clásicos caen dentro de la categoría de **sistemas de clave simétrica o secreta** (en contraposición a los de clave pública, que no serían inventados hasta 1976). En muchos casos el propio algoritmo de cifrado debía permanecer secreto y era parte de la clave.

En esencia, los **cifradores por transposición** funcionan permutando o intercambiando los caracteres del mensaje, reordenándolos así de una forma concreta. Un ejemplo sencillo sería considerar el mensaje en bloques de cuatro caracteres, y reordenar los caracteres de cada bloque siguiendo un patrón específico; por ejemplo, 4132, que se interpretaría de la siguiente forma: el cuarto carácter se intercambia con el primero, el segundo con el primero, el tercero queda invariante y, por último, el cuarto se cambia finalmente por el segundo.

Por otro lado, los **cifradores por sustitución** literalmente sustituyen cada carácter el mensaje original por otro, de acuerdo a un patrón concreto. En función de cómo se haga esta sustitución, hablaremos de **cifradores monoalfabéticos** o **polialfabéticos**. Veremos todos los detalles a continuación.

Pero comencemos por el principio: **¿qué es exactamente la criptografía?**

¿Qué es la criptografía?

La **criptografía** (del griego *krypto* y *logos*, que significa «el estudio de lo oculto, lo escondido») es la ciencia que trata sobre escribir mensajes que nadie pueda entender, a menos que tenga un elemento secreto, llamado **clave**.

Estrictamente hablado, la criptografía es sólo una de las ramas de un concepto mayor, la **criptología**, que engloba a la **criptografía** y al **criptoanálisis**. Ésta última trata de conseguir el objetivo contrario al primero: **atacar y descifrar los mensajes cifrados sin el conocimiento de la clave**.

Conviene aquí hacer una aclaración sobre un error muy común; tanto que puede verse escrito habitualmente en todo tipo de publicaciones (incluso académicas). Los términos correctos en español para referirse al proceso de protección son **«cifrar»** (y NO encriptar), y el texto resultante **«está cifrado»** (y no encriptado). Los elementos secretos utilizados en este proceso se denominan **«claves»** (y no llaves). *Encriptar*, *encriptado* y *llaves* son términos incorrectos (que provienen probablemente de una traducción directa del inglés), y que, al menos, te servirán para distinguir quién conoce el tema en profundidad y quién no.

2.2. Historia de la criptografía

La historia de la criptografía abarca muchos miles de años. Puede decirse que el periodo clásico de la misma se extiende desde sus orígenes, casi con el nacimiento de la escritura, hasta el final de la Primera Guerra Mundial, cuando aparecieron las primeras máquinas electromecánicas. Hasta ese momento todos los procesos se realizaban a mano, con lápiz y papel, o, como mucho, con sistemas mecánicos muy sencillos (veremos a continuación que un simple palo, por ejemplo).

En el periodo entre las dos grandes guerras mundiales, los sistemas de cifrado empiezan a sofisticarse y aparecen las primeras máquinas electro-mecánicas que, llegada la Segunda Guerra Mundial juegan un papel decisivo en su desarrollo. Veremos en nuestro estudio detallado de la *máquina Enigma* alemana, que algunos expertos sostienen que su criptoanálisis permitió acortar la duración de la guerra en varios años y salvar muchos miles de vidas. Patentada en 1918, hizo de los criptógrafos una pieza fundamental en el desarrollo del conflicto bélico.

Todo cambiaría, por supuesto, con la llegada de los microprocesadores y la era digital al final de los años setenta. La complejidad de los algoritmos y procedimientos eran ya completamente inabordables para los humanos. Al mismo tiempo, la criptografía comenzó a abandonar los entornos exclusivamente militares y diplomáticos, donde había residido hasta entonces. Esto se debió a dos hechos cruciales que abrieron la criptografía al mundo civil: la publicación por parte del NIST del *Data Encryption Standard* (el famoso algoritmo DES), y el otro, la invención de la criptografía de clave pública.

Hoy en día la criptografía es directamente dependiente de la capacidad de cómputo que nos aporta el hardware. Cifrados de no muchos meses, están «rotos» por criptosistemas. Aplicaciones de mensajería instantánea como WhatsApp, no basan sus actualizaciones periódicas en simples activaciones de tics a la lectura de mensajes o métodos de borrado, dichas actualizaciones tienen misiones mucho más importantes como la actualización de los algoritmos de cifrado, base de la credibilidad del secreto de las comunicaciones en este tipo de mensajería.

Los inicios de la criptografía clásica

La criptografía es una disciplina muy antigua, y sus orígenes se remontan prácticamente a los propios orígenes de la escritura. Se conocen de hecho, usos aún más antiguos, que se remontan a un conjunto de jeroglíficos tallados en un monumento egipcio del Imperio Antiguo. El escriba encargado de decorar una sala del mismo con una oda a las virtudes de su señor decidió utilizar un código sencillo de sustitución jeroglífica, cambiando algún que otro símbolo por uno de sus sinónimos menos conocidos. Este escriba no utilizó un sistema de cifrado propiamente dicho, pues únicamente cambió algunos jeroglíficos aquí y allá, sobre todo al final de los documentos.

¿Por qué lo hizo? Seguramente nunca lo sabremos, pero es posible que la intención fuera ocultar, ante el observador casual, el secreto de ciertos rituales religiosos, haciendo los textos difíciles de leer.

Sin embargo, los primeros ejemplos más serios llegarían, como tantas otras cosas, a través de sus usos militares. Se sabe que en la Grecia clásica, en el siglo quinto antes de nuestra era, se utilizaban ya métodos criptográficos para proteger mensajes e informaciones.

Un ejemplo de ello es la **escítala**, que se utilizó durante la guerra entre Atenas y Esparta. El historiador griego Plutarco la describe de la siguiente forma:

> «La escítala era un palo o bastón en el cual se enrollaba en espiral una tira de cuero. Sobre esa tira se escribía el mensaje en columnas paralelas al eje del palo. La tira desenrollada mostraba un texto sin relación aparente con el texto inicial, pero que podía leerse volviendo a enrollar la tira sobre un palo del mismo diámetro que el primero».

Según algunos documentos históricos, este sistema se utilizó por los gobernantes de Esparta para transmitir instrucciones secretas a los generales de su ejército, principalmente durante las campañas militares. Obviamente, para poder utilizar este método, tanto el emisor como el receptor del mensaje debían disponer de un bastón con el mismo grosor y longitud (que podrían considerarse la clave del sistema).

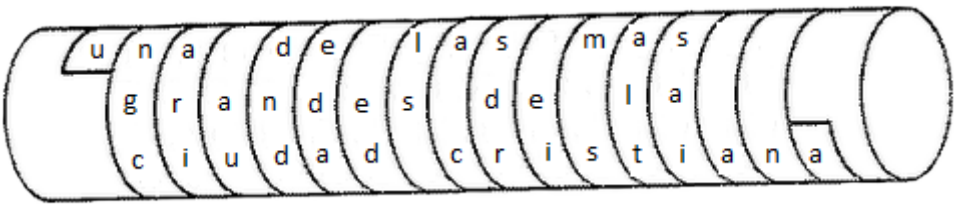


Figura 1. Escítala espartana con un mensaje codificado

Al desenrollar el mensaje del cilindro o bastón todas las letras del mismo siguen escritas en la cinta, pero su orden y posiciones relativas parecen no tener sentido, por lo que la lectura lineal del mensaje resulta incomprensible. Por ejemplo, en el caso de la **Figura 1**, la cinta codifica el mensaje:

M = «una de las más grandes de la ciudad cristiana»

Aunque a simple vista sólo puede leerse el criptograma:

C = «ungcariaudndedaedlsacsdreimsaltwaiana»

Para poder leer el mensaje, el receptor debía utilizar otro bastón del mismo tamaño y enrollar la cinta de papel escrita alrededor de él. Este sistema tenía la ventaja de que era rápido y bajo condiciones normales, estaba libre de errores, lo que lo hace muy útil en la presión de un campo de batalla. Sin embargo, incluso para la época, se caracteriza por un nivel de seguridad muy bajo y podía romperse con facilidad.

A pesar de ello, la importancia de este método de codificación llegó a ser tan grande que, según muchos historiadores, la famosa expresión «*ostentar el bastón de mando*», podría tener su origen en esta práctica, pues en él residía la seguridad de las comunicaciones y, por tanto, de la vida política de la antigua Grecia.

El cifrado de César

El imperio romano, como no podía ser de otro modo, también conocía e hizo uso de la criptografía. El ejemplo más conocido es, sin duda, el denominado **cifrado de César** y sus variaciones. Su nombre proviene, efectivamente, del emperador Julio César (100-44 a.C.), que la utilizó con un desplazamiento a la izquierda de tres posiciones para proteger sus mensajes de carácter militar.

El historiador Suetonius describía este método de la siguiente forma:

> *«Si él [César] tenía algo que decir, lo escribía en cifra, es decir, cambiando el orden de las letras del alfabeto, de modo que no se pudiese reconocer ni una palabra. Si alguien quiere descifrar eso, y obtener su significado, deberá sustituir la cuarta letra del alfabeto, es decir, la D, por una A, y así sucesivamente con todas las letras»*

Puesto que el número de posiciones a desplazar (que podríamos considerar la clave de este sistema) era siempre tres, los alfabetos a utilizar eran:

Original	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
Posición	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Cifrado	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

Obviamente, en esa época histórica, la letra «ñ» no existía. Para cifrar un texto en español que incluya esa letra, no hay más que añadirla a los alfabetos de texto en claro y cifrado, y adaptar consecuentemente el módulo de la suma.

Octavio Augusto, heredero político de Julio César, utilizaría un método similar, pero desplazando una sola posición el texto: «cuando escribía en cifra ponía la *b* por la *a*, *c* por *b* y así con las otras letras; por *x* ponía dos *a*».

Veamos un ejemplo de aplicación de este método. Si asignamos a cada letra un número (`A =00, B =01, C=02,..., Z=25`), y consideramos un alfabeto de 26 letras, la transformación criptográfica no es más que aplicar la siguiente expresión modular:

$$C = (M - 3) \ (mód\ 26)$$

Donde *C* es el correspondiente carácter del criptograma y *M* el del texto en claro. Por supuesto, esta expresión se puede generalizar para cualesquiera valores *k* y *n*:

$$C = (M + k) \ (mód\ n)$$

Donde ahora *k* es el desplazamiento a aplicar a cada letra (a la derecha positivo, a la izquierda negativo), y *n* la longitud del alfabeto que utiliza el mensaje a cifrar.

Tomaremos en nuestro caso como mensaje **M = FIRMA LA PAZ**. Sin más que ir aplicando la expresión anterior a cada letra, se obtiene el correspondiente criptograma. Para la primera letra **`F (F=5)`**:

$$C = (5 - 3) \ (mód\ 26) = 2$$

Que corresponde a la letra *C*. Y así sucesivamente hasta completar el mensaje. Como puedes comprobar fácilmente, el mensaje cifrado queda finalmente **`C=CFOJXIXMXW`**. Como veremos más adelante, este método se clasificaría como un cifrado del tipo de ***sustitución monoalfabética***.

Casi no haría falta decirlo, pero obviamente este método de cifrado (ni ninguno de los clásicos) NO es seguro hoy en día. La razón es clara: **existen únicamente 26** (o 27, según el lenguaje en el que esté el mensaje en claro y el alfabeto que se utilice) **claves diferentes**. Por tanto, basta con probarlas todas, cosa que puede hacerse a mano en unos minutos e instantáneamente con un ordenador.

Los sistemas que hemos presentado ilustran los dos principios esenciales en los que se basa la criptografía clásica: la **sustitución** y la **transposición**. El cifrado de César es un ejemplo de cifradores de sustitución (cada una de las letras del mensaje original tiene una correspondencia fija en el mensaje cifrado), mientras que la escítala espartana lo es de transposición (las letras simplemente cambian de sitio o se transponen, por tanto, las letras son las mismas en el mensaje original y en el cifrado). Veamos ahora ambos métodos con mayor detalle.

2.3. Cifradores de sustitución

Como acabamos de comentar, con solo unas pocas claves posibles, el cifrado de César es completamente inseguro, nada más que una curiosidad histórica hoy en día. Sin embargo, su estudio proporciona las bases para entender mejor los cifradores modernos, que, en esencia, están construidos sobre los mismos principios.

Para incrementar enormemente el espacio de claves, basta con permitir una sustitución arbitraria en ese esquema en lugar de una fija. Este simple cambio haría que pasásemos de **26 posibles alfabetos de cifrado** a cualquier permutación de ellos, lo que hace un total de **26 combinaciones**, o más de **$4 \cdot 10^{26}$ claves**.

Entonces, con esa cantidad de claves, **¿tenemos ya un cifrador seguro?** Pues tampoco, porque existe otra línea de ataque, de **criptoanálisis**, por el que caen todos los cifradores de sustitución clásicos. Este ataque se basa en el estudio de la frecuencia de aparición de las letras en el criptograma y compararla con la distribución de frecuencias de las letras en el idioma en el que está escrito el mensaje.

En efecto, todos los lenguajes humanos tienen una gran redundancia. Por ejemplo, en el lenguaje castellano considerando un alfabeto de 27 letras, podemos formar **tres grupos de frecuencias relativas**: uno de ***alta frecuencia***, otro de ***frecuencia media*** y un tercero de ***frecuencia baja***, como se muestra en la Figura 2.

Como puede observarse, las letras más frecuentes son la *E*, la *A* y la *S*. Para atacar un texto que suponemos cifrado con un cifrador de sustitución monoalfabético, procederíamos entonces de la siguiente forma:

1. Tendríamos que **hacer suposiciones sobre el lenguaje en el que está escrito el texto**, en función de donde fue capturado, el emisor o el destinatario (si se conocen), o cualquier otro dato que pueda darnos pistas en este aspecto. Si no, tendríamos que probar en los lenguajes más comunes, inglés, español, alemán, etc.
2. Realizaríamos un **estudio de las frecuencias relativas de cada letra en el criptograma**, formando así una tabla ordenada similar a la de la Figura 2.
3. **Compararíamos ambas listas** y comenzaríamos haciendo asignaciones tentativas de letras y formando un «esqueleto» provisional del mensaje, comprobando si cobra sentido.
4. Otros enfoques más sistemáticos consisten en **buscar regularidades en el criptograma**. Por ejemplo, encontrar secuencias de letras que se repiten y tratar de deducir sus equivalentes en claro (las palabras *THE*, *AND*, *OF* en inglés, o *DE*, *EN*, *PARA* en castellano).

Otra herramienta poderosa de criptoanálisis es estudiar la frecuencia de las combinaciones de dos letras, conocidos como **digramas**; o de *n* letras, conocidos como ***n*-gramas**. En general, para textos «genéricos» se observa que los tres digramas con mayor frecuencia relativa en castellano son `DE`, `ES` y `EN`, a los que les siguen `LA`, `OS`, `AR`, `UE`, `RA`, `RE`, `ER`, `AS`, `ON`, `ST` y `AD`, en orden decreciente. Por otro lado, los trigramas más comunes son, de nuevo de mayor a menor frecuencia de aparición: `QUE`, `EST`, `ARA`, `ADO`, `DEL` y `CIO`.

Frecuencia Alta		Frecuencia Media		Frecuencia Baja	
E	13,11	C	4,85	Y	0,79
A	10,60	L	4,42	Q	0,74
S	8,47	U	4,34	H	0,60
O	8,23	M	3,11	Z	0,26
I	7,16	P	2,71	J	0,25
N	7,14	G	1,40	X	0,15
R	6,95	B	1,16	W	0,12
D	5,87	F	1,13	K	0,11
T	5,40	V	0,82	Ñ	0,10

Figura 2. Distribución de frecuencias del idioma castellano, en porcentaje

Asimismo, existirán **digramas con frecuencia nula** como sería el caso de *QA*, *KK*, *ÑL*, *WZ*, etc., pues no forman parte de palabra alguna ni son término e inicio de dos palabras contiguas. En inglés, el digrama más común es *TH*.

Veamos ahora cómo podemos utilizar más de un alfabeto para mejorar nuestros cifradores de sustitución.

Cifradores de sustitución polialfabéticos

Ya hemos visto que en el cifrado César la letra *D* del texto en claro se cifraba *siempre* como la letra *A*, lo que lo convertía en un cifrador monoalfabético. Supongamos ahora que deseamos diseñar un algoritmo mejorado, de forma que, a los caracteres impares, por ejemplo, aplicamos un desplazamiento a la derecha del alfabeto y a los caracteres pares otro a la izquierda (o cualquier otra combinación que se nos ocurra).

¿Qué gran ventaja tiene este enfoque? Pues que el criptograma que se produce ahora ya no existe una correspondencia única entre los caracteres del texto en claro y los de un alfabeto de cifrado: una letra concreta se cifrará como un carácter u otro, dependiendo de si se encuentra en una posición par o impar en el texto en claro.

Al utilizar dos desplazamientos diferentes, tenemos que utilizar necesariamente dos alfabetos de cifrado también diferentes, por lo que algunos caracteres, dependiendo de su posición, se cifrarán como dos caracteres distintos.

Este procedimiento de cifrado será el origen de, entre otros, al cifrador de Vigenère, que analizaremos a continuación. En función del número de alfabetos utilizados, la sustitución polialfabética se clasifica en ***periódica***, como en el ejemplo anterior, cuyo período es dos, o ***no periódica***, si la clave utilizada es al menos tan larga como el mensaje.

Cifrador de Vigenère

La mayoría de los cifradores polialfabéticos son, en realidad, periódicos, y dicho período vendrá determinado por la longitud de la clave de cifrado. La única excepción a esta regla son los no periódicos, también llamados cifradores de clave continua. Su ejemplo más característico es el **cifrador de Vernam**, que posee una clave de igual longitud (o mayor) a la del texto en claro.

Entrados ya en el siglo XVI encontramos el ejemplo del cifrador polialfabético más conocido, de la mano de Blaise de Vigenère, un diplomático y criptólogo francés del siglo XVI. El **cifrado de Vigenère** se denomina así porque incorrectamente se atribuyó a él su invención en el siglo XIX, cuando en realidad, fue ideado por Battista Bellaso, un noble veneciano y contemporáneo de Vigenère, alrededor de 1550.
El sistema funciona sumando a cada carácter del texto en claro uno de los caracteres de la clave, que va rotando durante el proceso de cifrado. Cuando la clave se «agota», vuelve a escribirse de forma cíclica bajo el mensaje:

Texto:	TOBEORNOTBETHATISTHEQUESTION ...
Clave:	RUNRUNRUNRUNRUNRUNRUNRU...
Criptograma:	KIOVIEEIGKIOVNURNVJNUVKHVMGZIA

En nuestro ejemplo, nuestro texto en claro es la famosa frase pronunciada por Hamlet, «To be or not to be, that is the question». La clave es *RUN*, que tendrá, obviamente, un periodo de *3*. Así, el primer carácter del texto en claro (T) se le aplica un desplazamiento equivalente al primer carácter de la clave (R), dando lugar a **T + R = (19 + 17) mod 26 = 10 = *K***; el segundo carácter (O) se cifra sumándolo con el segundo carácter de la clave (U), **O + U = (14 + 20) mod 26 = 8 = I**, etc. El resultado final será el criptograma **C = KIOVIEEIGKIOVNURNVJNUVKHVMGZIA**.

Es importante observar que letras repetidas del texto en claro se cifran habitualmente de forma distinta, en función de su posición respecto a la clave. También ocurre que una misma letra del criptograma puede ser originada por letras diferentes del texto en claro.

Para acelerar este proceso de cifrado, se han utilizado varias técnicas diferentes a lo largo de los años, desde ruedas de cifrado (que, a pesar de ser útiles, era necesario un cifrado rápido) hasta tablas impresas. Estas tablas se denominan ***tablas de Vigenère***, y un ejemplo se muestra en la **Figura 3**.

Para cifrar un texto utilizando una tabla de Vigenère comenzamos buscando la primera letra del mensaje a cifrar en la primera fila de la tabla, y haciendo lo mismo con la primera letra de la clave en la primera columna. Es decir, ***las letras del mensaje se buscan en las filas y las de la clave en las columnas.***

Una vez encontrados, sólo debemos buscar la letra que se encuentra en la intersección de la fila y columna correspondientes. De esta forma, la letra *P* cifrada con la clave *E* generará la letra *T* en el criptograma. Veamos un ejemplo completo.

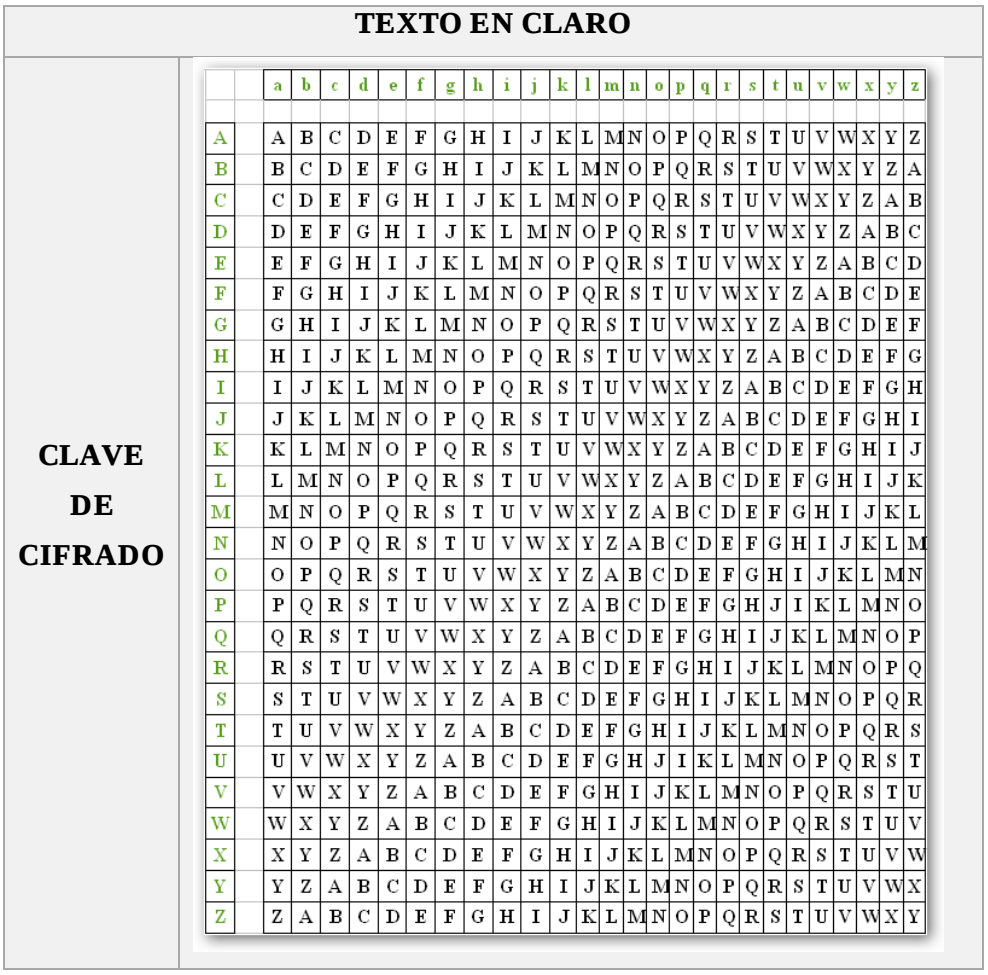


Figura 3. Tabla de cifrado y descifrado de Vigenère

Ejemplo

Utilizando la *tabla de Vigenère* y la clave **K = UNIR**, cifra el siguiente mensaje: **M = ME ENCANTA ESTA ASIGNATURA**.

Solución: Como hemos visto, podemos resolver este problema escribiendo la clave de forma cíclica bajo el mensaje a cifrar y sumando los valores numéricos correspondientes a las letras:

M E E N C A N T A E S T A A S I G N A T U R A
U N I R U N I R U N I R U N I R U N I R U N I...

De esta forma, el primer carácter del criptograma sería $C_1=(M+U)=(12 + 20) \bmod 26 = 32 \bmod 26 = 6 = G$. También podemos hacer uso directamente de la tabla, y en la intersección de las columnas correspondientes a la *M* y a la *U* encontramos, efectivamente la *G*. El criptograma final queda:

C = GRMEW NVKUR AKUNA ZAAIK OEI

Formalmente, si denominamos *d* a la longitud de la clave $K = k_1...k_d$, donde k_i $(i = 1,..., d)$ indica el desplazamiento del alfabeto i-ésimo, la función de transformación de Vigenère para cifrar viene dada por:

$$C_i = E_{ki}(M_i) = (M_i + k_i) \bmod n$$

Donde de nuevo *n* es la longitud del alfabeto del mensaje de entrada.

Para llevar a cabo el descifrado de un mensaje haciendo uso de una tabla de Vigenère, no hay más que proceder de forma inversa. Ahora los papeles de filas y columnas se intercambian respecto al cifrado:

1. Comenzamos buscando en la fila de *Texto en claro* el primer carácter de la clave, y bajamos ahora por su columna hasta encontrar el primer carácter del criptograma.
2. En ese momento, nos movemos hacia la izquierda, hasta llegar a la columna de *Clave de cifrado*, donde encontraremos ya la letra correspondiente del mensaje descifrado.

Por ejemplo, descifremos las primeras letras del ejemplo anterior. El primer carácter de la clave es *U*, por lo que la buscamos en la fila etiquetada como *texto en claro*. Ahora bajamos por esa columna hasta encontrar la primera letra del criptograma, G, en este caso. Moviéndonos hasta la columna de *clave de cifrado*, encontramos el correspondiente carácter del *texto en claro*, *M*, que comprobamos que es correcto. Solo queda repetir el procedimiento para el resto del mensaje.

A pesar de su simplicidad, el cifrado de Vigenère fue inexpugnable durante mucho tiempo. Muchas personas trabajaron en su criptoanálisis, desde el famoso Casanova hasta el pionero de la computación Charles Babbage. Sin embargo, la primera solución correcta fue publicada en 1863, algo más de trescientos años después de su invención, por Friederich Kasiski, un oficial de infantería prusiano.

En esencia, su criptoanálisis consiste en tratar de encontrar el período del cifrador (la longitud de la clave), de forma que podamos descomponer el cifrador polialfabético en *n* cifradores monoalfabéticos, que ya sabemos cómo atacar.

Existen dos métodos básicos para conseguir este objetivo: el ***método de Kasiski*** y el ***índice de coincidencia (IC)***. Veamos a continuación más detalles sobre ellos.

El método de Kasiski

El **método de Kasiski** consiste en un método de criptoanálisis de cifradores polialfabéticos periódicos, que analiza repeticiones en el texto cifrado «*con el fin de determinar el periodo (longitud de la clave) que se usó para cifrarlo*».

Lo estudiaremos a través del ejemplo de la famosa frase de Hamlet, para facilitar su comprensión. Recuerda que el mensaje, clave y criptogramas concretos son:

TEXTO:	TOBEORNOTBETHATISTHEQUESTION...
CLAVE:	RUNRUNRUNRUNRUNRUNRUNRU...
CRIPTOGRAMA:	KIOVIEEIGKIOVNURNVJNUVKHVMGZIA

Kasiski se dio cuenta de que, para un mensaje suficientemente largo, es habitual encontrar patrones que se repiten en el criptograma. Cuando la longitud de estos patrones es mayor o igual que 3, es probable que esas palabras sean también iguales en el *texto en claro*.

Esto significa que han sido cifrados con la misma sustitución, por lo que la distancia entre las posiciones de comienzo de las palabras iguales será un múltiplo del periodo que se usó para cifrar el texto.

En nuestro ejemplo hemos subrayado el patrón (4-grama) «KIOV», que se encuentra repetido a una distancia de 9 caracteres, y también «UN» a 6. Esto implica que el periodo que se usó para cifrar el texto debe ser divisor de 9 y 6, lo que sólo deja al 3 como opción. En nuestro caso la clave es «RUN» que es, efectivamente, de tres letras, por lo que el ***método de Kasiski*** ha funcionado correctamente en este caso.

Podemos inferir aún más datos: como la clave es de longitud 3, sabemos que las letras en primera, cuarta, séptima,... posición están cifradas bajo el mismo carácter de la clave, aunque aún no sepamos cuál es. Lo mismo aplica para las posiciones segunda, quinta, octava, etc. y tercera, sexta, novena, etc., cifradas cada una de estas series con otra letra diferente. De esta forma, podemos separar el ***cifrado polialfabético*** original en ***n cifrados monoalfabéticos simples***, donde n es la longitud de la clave. En cada uno de ellos podemos aplicar las técnicas de análisis de frecuencia que vimos en la sección anterior.

En cualquier caso, es importante entender que este método no es infalible. Puede ocurrir que encontremos patrones iguales en el criptograma cuya separación no sea múltiplo de la clave o, incluso, que sea un número primo. En estos casos, este método no nos sería de ayuda, y tendríamos que buscar otra aproximación.

En la práctica, para evitar en lo posible estas pistas falsas, se debe procurar contar con un criptograma de muchos caracteres (idealmente varias centenas), para poder buscar coincidencias largas, de una longitud mayor o igual a 3 y que se repitan más de una vez.

Índice de Coincidencia

Otro método para averiguar el período de un cifrador polialfabético es el denominado **índice de coincidencia**, propuesto por William F. Friedman en 1922. Su libro *El índice de coincidencia y sus aplicaciones en criptografía* (1922) fue considerada durante años una de las mayores contribuciones a la criptología, debido principalmente al tratamiento matemático riguroso y al enfoque estadístico que, por primera vez, se aplica a la criptología.

Este concepto fue básico, de hecho, en el posterior criptoanálisis de las máquinas de rotores, como *la Enigma*, que tan importantes fueron en el desarrollo de la Segunda Guerra Mundial.

Aunque no estudiaremos aquí la derivación del resultado, el índice de coincidencia puede definirse como la probabilidad de que dos letras de un texto cifrado elegidas al azar sean la misma, y se calcula como:

$$IC = \frac{\sum_{i=0}^{n-1} f_i(f_i - 1)}{n(n-1)}$$

Donde n es la longitud del texto cifrado y f_i la cantidad de veces que aparece la letra i-ésima en dicho texto. A partir de esta expresión y del estudio de la distribución de frecuencias de aparición de letras en nuestro idioma, puede construirse una tabla como la siguiente:

d	IC
1	0,072
2	0,054
3	0,049
4	0,046
5	0,044
6	0,040
...	
Grande	**0,037**

Figura 4. Índice de coincidencia para cifras con período *d*

Con esta tabla, sólo falta calcular el IC del texto cifrado que queremos romper y buscar el valor obtenido en la tabla. Si está, por ejemplo, alrededor del valor **`IC = 0,049`** significa que estamos ante un **cifrador polialfabético de (posiblemente) periodo 3.** Por otro lado, si obtenemos un valor `IC = 0,072` este método nos dice que se trataría en realidad de un **cifrador monoalfabético (polialfabético de periodo 1)**, por lo que el IC sirve, en primer lugar, para determinar si se trata de un cifrado mono o polialfabético y, en segundo lugar, para estimar el periodo del mismo en el segundo caso. Veamos un ejemplo de aplicación.

Ejemplo:

Haciendo uso del IC, determina si el siguiente criptograma pertenece a un cifrado por sustitución monoalfabética o polialfabética.

C = GUVQA EQORN GVCOG PVGUG IWTQF GSWGG NRTKO GTCNW OPQSW GFGUE KHTGG UVGOG PUCLG ANQRW DNKSW GGPGN HQTQF GNCCU KIPCV WTCVG PFTCO CURWP VQUGP UWPQV C

Solución: aplicando la ecuación correspondiente a los 121 caracteres del criptograma se obtiene: IC = 0,071. Como este valor se aproxima mucho a 0,072, podemos concluir que se trata de un cifrado de tipo monoalfabético. ¿Podrías obtener el texto *en claro*? ¡Tiene premio!

Criptoanálisis clásico: resumen

En resumen, el **procedimiento más habitual para analizar un criptograma supuestamente cifrado con un cifrador polialfabético** es el siguiente:

1. **Análisis de la distribución de frecuencias de los distintos caracteres del criptograma.** Si es similar a la del lenguaje en el que se cree está escrito el mensaje, el cifrado será monoalfabético; en caso contrario, será polialfabético.
2. **Aplicación del método de Kasiski** para corroborar que el criptograma fue obtenido con un cifrador polialfabético y en ese caso, obtener una primera estimación del periodo del cifrador.
3. Tanto si el método anterior falla como si no, conviene **calcular en cualquier caso también el índice de coincidencia del texto** para comprobar si se obtienen los mismos resultados.

2.4. Caso de estudio: la *máquina Enigma*

A principios del siglo XX comienzan a desarrollarse diversos dispositivos puramente mecánicos de cifrado, que hacían uso de **rotores o ruedas** para realizar un cifrado polialfabético. Esto daba lugar a un número muy alto, varios millones, de posibles alfabetos.

Dado que el criptoanálisis aún se encontraba en pañales y que las técnicas de ataques disponibles eran totalmente manuales y poco más sofisticadas que las que hemos estudiado en este capítulo, estas máquinas fueron inexpugnables durante varios años.

De estas máquinas, que se utilizaron sobre todo para enviar mensajes cifrados durante la Segunda Guerra Mundial, destaca por sus poderosas características y todo el misterio y secretismo que la rodeaba la llamada ***máquina Enigma***.

De aspecto físico era muy similar al de la **Figura 4**, esta máquina fue inventada por el ingeniero alemán Arthur Scherbius en el año 1923 y disponía de un conjunto de rotores (o ruedas dentadas) montados sobre un eje, alrededor del que había 26 contactos eléctricos, uno por cada letra del alfabeto inglés. Disponía además de un teclado similar al de la máquina de escribir, por lo que, al pulsar cada letra, aparecía otra en una consola y que simultáneamente, y según las versiones, iba escribiendo en una tira de papel que iba saliendo por un lateral.

Los rotores se desplazaban secuencialmente; es decir, al cifrar un carácter el primer rotor avanzaba una posición y sólo cuando éste había realizado una rotación completa, el segundo se desplazaba un carácter, y así sucesivamente. A pesar de su simplicidad mecánica, estas máquinas producían un cifrado polialfabético fuera del alcance del criptoanálisis manual de la época. Por ejemplo, en un sistema con 4 rotores, se desarrollaban un total de 456.976 alfabetos. Si aumentamos los rotores a 5, esta asciende a una cantidad que ya es considerable, de 11.881.376. Se sabe, además, que existieron versiones especiales de la máquina con 10 rotores, para uso exclusivo de los mensajes personales enviados por Hitler.

Figura 3. Máquina Enigma de cuatro rotores. (Fuente: www.u-historia.com)

Para descifrar el mensaje, era necesario que el receptor dispusiera de otra ***máquina Enigma*** **y una copia del libro de códigos** que contenía la posición inicial de los rotores para ese día. Obviamente una regla de oro para cualquier militar alemán en este periodo era que nunca debe permitirse que la máquina y el libro de códigos cayeran en manos enemigas. Se cuenta, de hecho, que los oficiales de los submarinos alemanes tenían órdenes específicas en este aspecto: la *máquina Enigma* y el libro de códigos era el material más valioso que iba a bordo del submarino, por encima de la vida de cualquier soldado y la suyas propias. Si debían elegir entre salvar sus vidas o la *máquina Enigma*, la máquina era prioritaria.

El ejército aliado, especialmente los británicos, establecieron un grupo de trabajo destinado específicamente a romper el cifrado Enigma en Betchley Park, en Inglaterra. Al frente del dicho grupo se encontraba el famoso Alan Turing, brillante matemático, considerado también el padre de la informática moderna.

Este equipo de élite consiguió, por fin, tras varios años de trabajo, romper el cifrado Enigma. Por supuesto, éste hecho se mantuvo en secreto, pues los alemanes cambiarían el sistema a la mínima sospecha y se perdería el factor sorpresa.

Así pues, los aliados esperaron una ocasión propicia para tratar de dar un golpe definitivo a la guerra. Dicha ocasión se presentó finalmente el 1 de junio de 1944, cuando los aliados interceptaron y pudieron descifrar un mensaje crucial: Hitler y su Alto Mando esperaban un ataque aliado masivo en Calais. Este conocimiento fue decisivo para que el general Eisenhower tomara la decisión de desembarcar sus tropas el 6 de junio en las playas de Normandia, para aprovechar el efecto sorpresa y multiplicar, así, el efecto del ataque sobre el ejército alemán. Este hecho supuso, según un artículo de The Guardian de 1995, un acortamiento de la guerra de por lo menos dos años y la salvación de miles de vidas.

Lo + recomendado

No dejes de leer...

Breve estudio sobre la historia de la criptografía

En este estudio del Instituto Nacional de Ciberseguridad (INCIBE) encontrarás un breve resumen de la historia de la criptografía, como un repaso de todas las cuestiones que hemos abordado en este tema, incluido ciberespionaje y criptobulos.

Accede al documento a través del aula virtual o desde la siguiente dirección web:
https://www.incibe-cert.es/blog/ciberespionaje-criptografia

Proyecto M4

El Proyecto M4 es un esfuerzo de computación distribuida dirigido a descifrar tres mensajes cifrados con la máquina alemana Enigma en su variante de cuatro rotores (M4, de ahí el nombre del proyecto), que fueron interceptados en el Atlántico Norte en 1942 y desde entonces permanecen presuntamente sin descifrar. Desde la página del proyecto pueden descargarse clientes *open source* para Unix y Windows.

Accede al documento a través del aula virtual o desde la siguiente dirección web:
https://www.youtube.com/watch?v=g5LZvytKrys

No dejes de ver...

Enigma

Título original: Enigma
Año: 2001
Duración: 119 min
País: Reino Unido y EE.UU.
Director: Michael Apted
Interpretación: Dougray Scott, Kate Winslet, Saffron Burrows

Una historia tan novelesca e importante para la Historia como la de la *máquina Enigma* no ha pasado desapercibida para el cine. En esta película, que protagoniza Kate Winslet, se cuenta la intrahistoria de unos operadores de Bletchley Park. El rigor histórico no es el mejor pero, al fin y al cabo, es una película cuyo fin último es entretener. No te la pierdas y te dará una aproximación a los actores humanos de esta parte de la Historia.

Secretos de la Segunda Guerra Mundial

Este interesante documental hace un exhaustivo recorrido por la historia de la *máquina Enigma* y el papel que jugó en el desarrollo y desenlace de la Segunda Guerra Mundial. Es un documento interesante para profundizar más en este tema.

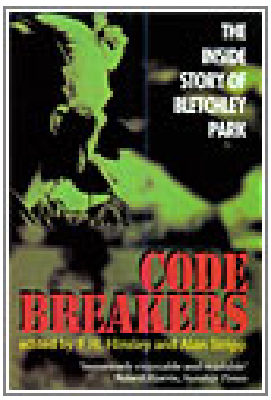
Accede al video a través del aula virtual o desde la siguiente dirección web:
https://www.youtube.com/watch?v=O9jqjbp1nxg

+ Información

A fondo

Codebreakers: The Inside Story of Bletchley Park

Hinsley, F.H. & Stripp, A. (1993). *Codebreakers: The Inside Story of Bletchley Park.* Oxford University Press. ISBN: 019285304X.

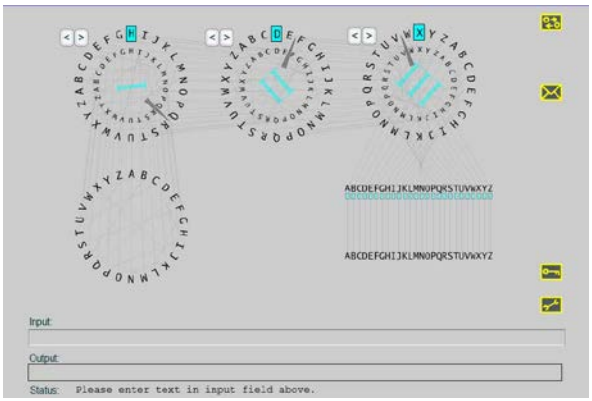


Este libro narra, en primera persona, la historia de Bletchley Park, el lugar en que se rompieron e interpretaron las transmisiones alemanas, italianas y japonesas durante la Segunda Guerra Mundial. Los protagonistas responsables de Ultra, el nombre en clave que los británicos dieron a todas las transmisiones de inteligencia de los enemigos del bando aliado, cuentan cuál fue su importancia y cómo se descifraron sistemas criptográficos como los empleados en la *máquina Enigma* alemana. El libro es una recopilación de relatos de los trabajadores de Bletchley Park, algunos bien conocidos en el mundo de la criptología, otros, héroes anónimos. Bletchley Park llegó a romper la criptografía de 4.000 mensajes alemanes al día y desarrollar las «bombas» lógicas, Colossus y Mark, precursores de los actuales ordenadores, con el único objetivo de romper códigos secretos. Como vino a decir Winston Churchill sobre Bletchley Park, «*esas instalaciones y la gente que allí trabajó fueron el arma secreta aliada que permitió ganar la guerra*». Nada más y nada menos.

Recursos externos

Simulador interactivo de la máquina Enigma

Este interesante enlace dispone de simulador online de una *máquina Enigma* de tres rotores. Permite configurar las claves, los rotores y sus posiciones iniciales y teclear texto que es automáticamente cifrado (o descifrado) mostrándose de forma visual qué camino recorre con cada letra. Además incluye opciones para enviar el texto cifrado a alguien por correo y es, en general, visualmente muy atractivo.



Accede al simulador a través del aula virtual o desde la siguiente dirección web:
http://enigmaco.de/enigma/enigma.swf

Test

1. Los dos grandes tipos de cifradores clásicos son:

A. De sustitución y transposición
B. Mono y polialfabéticos
C. Mono y poligrámicos
D. El cifrado de César y el Vigenère

2. El siguiente criptograma ha sido interceptado en la organización donde trabajas `C=VLOHHVHVWRWXUHVSXHVWDVHUDFRUUHFWD`. Se sabe también que ha sido cifrado mediante un cifrador de sustitución monoalfábetica de clave +3 (es decir, avanza cada carácter 3 posiciones a la derecha). ¿Cuál es el texto en claro correspondiente?

A. `LO SIENTO TU RESPUESTA SERA INCORRECTA`
B. `SI LEES ESTO TU RESPUESTA SERA CORRECTA`
C. `LA POSICION DEL ENEMIGO ES DESCONOCIDA`
D. `NECESITAMOS AYUDA URGENTE POR FAVOR`

3. La técnica básica de criptoanálisis de textos cifrados con cifradores de sustitución monoalfábeticos es:

A. Índice de coincidencia
B. Método de Kasiski
C. Análisis de frecuencia
D. Análisis de digramas

4. Las dos letras más frecuentes en el idioma español son:

A. La `A` y la `S`
B. La `A` y la `O`
C. La `A` y la `E`
D. La `E` y la `O`

5. El cifrado de Vigenère es un tipo de cifrador:

A. Polialfabético
B. Monoalfabético
C. Es una variante del cifrado César
D. De transposición

6. Ante una situación de emergencia, y al no disponer de ningún ordenador cerca, decides dejar un mensaje cifrado para un amigo. El texto en claro es el siguiente `T=NOS VEMOS EL MARTES A LAS DOCE DONDE SIEMPRE`, y decides utilizar el método de Vigenère, que puede ser calculado fácilmente a mano, con la clave `K = ASIMOV`. ¿Cuál sería el criptograma que debes entregar a tu amigo?

A. `OHBATIPLNYBWSMNFPAJLMBRAEHWQTOJXVCGA`

B. `OHBITIPLNYBWSMNFPHBLMBRAEHWQTOJXVCGA`

C. `AEHWQTOJXVCGAJKFPHBLMBRAEHWQTOJXVCGA`

D. `FPHBLMBRAEHWMNFPHBLMBRAEHWQTOJXVCGA`

7. Ponte ahora en el lugar de tu amigo, que recibe el mensaje, que conoce el método de cifrado pero no la clave con la que fue cifrado. ¿Qué procedimiento podrías utilizar para obtener una primera aproximación a la longitud de la clave utilizada en el cifrado anterior?

A. Índice de coincidencia

B. Método de Kasiski

C. Análisis de frecuencia

D. Fuerza bruta

8. Has interceptado un mensaje cifrado con un cifrador clásico. De momento no tienes más datos, por lo que decides calcular el índice de coincidencia del texto. El resultado resulta ser de 0,056. ¿Qué te dice este valor?

A. Que el texto ha sido, posiblemente, cifrado con un cifrador polialfabético de periodo 2 (es decir, con dos alfabetos)

B. Que el texto ha sido, posiblemente, cifrado con un cifrador polialfabético de periodo 3 (es decir, con tres alfabetos)

C. Que el texto ha sido, posiblemente, cifrado con un cifrador monoalfabético

D. Que el texto ha sido cifrado, con todo seguridad, con un cifrador monoalfabético

9. En tu nuevo trabajo como criptoanalista, recibes tu primer encargo. Se trata de un criptograma, `C = AOWWOROSXZWSHSGWOQIZWROSXZWAOSPDIGGEGLQ`, del que sabes a ciencia cierta que fue cifrado con un esquema clásico. De acuerdo al método de Kasiski, ¿cuáles son los posibles valores para el periodo (número de alfabetos) del criptograma?

A. `1, 2, 4, 8`
B. `1, 2, 3, 4, 6, 12`
C. `1, 2, 8, 16`
D. `1, 2, 8`

10. El criptoanálisis por parte del bando aliado de la máquina Enigma jugó un papel importante en el desenlace de la Segunda Guerra Mundial. ¿Qué matemático inglés fue su principal artífice?

A. Winston Churchill
B. Arthur Scherbius
C. Alan Turing
D. Kurt Gödel