RAPPORT D'INFORMATIQUE

Cryptographie – Cryptanalyse – Interface graphique

CACHARD Benoit

PENOT Baptiste

ACETO Lilian

ARNAUD Théo

Elèves-Ingénieurs à l'E.N.S.G.

Novembre 2021 - Décembre 2021

Notice analytique

Titre du document :	Cryptographie – Cryptanalyse – Interface graphique
Cadre et nature du travail :	Programmation - Python
Date de début et de fin :	Novembre 2021 – Décembre 2021
Date de publication :	18/12/21
Auteurs :	CACHARD Benoit PENOT Baptiste ACETO Lilian ARNAUD Théo
Encadrement :	BAVILLE Paul FAY-VARNIER Christine
Résumé :	Le rapport présente la cryptographie et la cryptanalyse en s'appuyant sur différentes méthodes en exemple, de différents types. Des détails sur la façon dont les méthodes et l'interface ont été codés sont également donnés. *This report presents cryptography and cryptanalysis by explaining different kinds and methods as examples. Details on the way used to code these methods and interface are given too.*
Mots-clés :	Cryptographie, Cryptanalyse, Interface Graphique, Codage, Décodage, Méthodes, clé *Cryptography, cryptanalysis, graphic interface, coding, decoding, key methods*
Nombre de pages :	18

Table des matières

Table des illustrations

I. Introduction

Ce projet d'informatique s'intègre dans le cadre de l'enseignement du Semestre 7 de l'Ecole Nationale Supérieure de Géologie de Nancy. Celui-ci constitue une bibliographie ainsi qu'une notice explicative du programme fourni.

L'objectif de ce rapport est de présenter diverses méthodes de Cryptographie via le langage Python. Le rendu final du projet se présente sous forme d'une interface graphique reliée aux diverses méthodes dans le but de faciliter l'interaction avec l'utilisateur. On expliquera également les méthodes à appliquer pour effectuer une cryptanalyse.

Dans un premier temps, nous présenterons le sujet plus en détail ainsi que les différentes méthodes utilisées. Ensuite nous nous attarderons sur la partie programmation pure du projet ainsi que sur les choix de présentation qui ont été faits. Enfin nous comparerons entre elles diverses méthodes de notre dossier final avant de conclure.

II. Présentation du sujet et des méthodes utilisées

Le sujet

Le besoin de générer des messages sécurisés compréhensibles uniquement par le créateur et le destinataire de ce dernier est apparu très tôt dans l'histoire. On peut facilement en juger par la quantité de méthodes existantes et l'ancienneté de certaines d'entre elles. L'action de coder un message s'appelle la **cryptographie**. Le fait d'essayer de comprendre par quelle méthode le message est codé et de le décoder est la **cryptanalyse**.

Les méthodes sont réparties en plusieurs grandes catégories :

- Les méthodes de **substitution mono-alphabétique** : consistent à utiliser un alphabet mélangé (avec les lettres dans un ordre inhabituel) et à remplacer les lettres de l'alphabet normal par ce dernier.
- Les méthodes de **substitution poly-alphabétique** : le chiffrement par substitution poly-alphabétique consiste à encrypter un message clair en substituant les lettres de celui-ci par d'autres, dans un procédé déterminé par la clef, et non un procédé fixe comme la méthode de substitution mono-alphabétique. La machine allemande Enigma, utilisée durant la Seconde Guerre Mondiale est la plus célèbre utilisation de cette technique. La substitution poly-alphabétique repose donc sur la possession d'une clef pouvant aussi bien être une suite de caractères qu'une suite de chiffres. Notons que pour chiffrer un message, la clef doit être répétée ou coupée pour avoir la même taille que le message clair.
- Les méthodes de **chiffres polygraphiques** : Les chiffres polygraphiques sont des chiffres pour lesquels on partage d'abord le message chiffré en groupes d'un certain nombre de lettres. Pour chacun de ces groupes, on opère alors un algorithme de chiffrement (le plus souvent, une substitution) pour chiffrer le message.

Une autre partie importante du projet est la mise en place d'une **interface graphique** grâce à la bibliothèque PyQt5. On utilise pour cela de nombreuses classes de PyQt comme celles permettant la création d'un widget, élément de base de l'interface graphique.

Cette interface joue un rôle central dans le rendu final du projet car elle permet de faciliter l'interaction entre l'utilisateur et le code, mais aussi de relier toutes les méthodes codées précédemment entre elles.

Les méthodes de Cryptographie

La substitution mono-alphabétique

- Méthode de César :

Cette méthode est probablement parmi les plus simples. Elle nécessite un mot d'entrée que l'on souhaite coder ainsi qu'une clé. Un nouvel alphabet décalé de n rang est généré, « n » étant la clé donnée par l'utilisateur. Chaque lettre du mot à coder est ensuite remplacer par la lettre associée dans le nouvel alphabet. Par exemple, pour une clé valant 4, un A est codé par un E.

- Méthode de Atbash :

Il s'agit d'une variante de la méthode de césar. Dans ce cas, pas besoin de clé, le nouvel alphabet est simplement l'alphabet à l'envers. Encore une fois, chaque lettre du mot à coder est remplacée par la lettre associée dans le nouvel alphabet (celui à l'envers). Par exemple, un A est codé par un Z et un C est codé par un X dans le cas de l'utilisation de la méthode de Atbash.

- Méthode du carré de Polybe :

Dans ce cas, le nouvel alphabet est en fait une matrice 5x5 dont chacune des cases contient une lettre, les lettres I et J sont affectées dans la même case et ne seront donc pas différenciées lors du décodage d'un message par cette méthode. Pour coder un message par cette méthode, on affecte chaque lettre du message à coder à ses coordonnées (x, y) dans la matrice 5x5. Les coordonnées x comme y sont donc comprises entre 0 et 4. Par exemple, le message OUI ET NON sera codé par la suite d'entiers suivante : 233413 0433 222322.

- Méthode du chiffre de Che :

Cette méthode mythique fut notamment utilisée par Che Guevara pour communiquer avec Fidel Castro. Elle est plus complexe que les précédentes car elle présente des étapes supplémentaires. Dans un premier temps, on effectue une simple substitution des différentes lettres du message à coder par les valeurs associées qui sont représentées dans le tableau ci-dessous :

A	B	C	D	E	F	G	H	I	J	K	L	M
6	38	32	4	8	30	36	34	39	31	78	72	70
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
76	9	79	71	58	2	0	52	50	56	54	1	59

38976 [clair]
25638 [clé]

53504 [crypté]

Figure 1 : tableau de substitution et principe du code de Che

On obtient donc une suite de chiffres compris entre 0 et 9. On regroupe cette suite en blocs de 5 chiffres que l'on additionne ensuite en modulo 10 avec une clé à 5 chiffres choisie par l'utilisateur. Cette action est illustrée à droite du tableau ci-dessus. Le message est alors crypté.

La substitution poly-alphabétique

- Méthode du chiffre de Vigenère

Toute la méthode repose sur l'utilisation d'une table : la table de Trithème, reprise par Vigenère.

Il s'agit d'un abaque sous la forme d'une matrice, comprenant 27 lignes et 27 colonnes. La première ligne et la première colonne servant de graduation, cette table permet de trouver un caractère en connaissant le message clair et sa clef.

Cette méthode est très efficace car l'analyse de fréquences classique n'est plus applicable dans ce cas puisque le remplacement d'une lettre par une autre est déterminée sur l'état de cryptographie du message, contrairement à la méthode mono-alphabétique.

La répartition des proportions ne permet donc plus d'obtenir un indice sur les lettres du message.

La méthode est appelée « méthode du masque jetable » lorsque la clef est aussi longue que le message clair et que certaines précautions ont été prises concernant celle-ci (caractères aléatoires …). Voici ci-dessous la table utilisée dans la méthode de Vigenère, tirée du site apprendre en ligne:

	Lettre en clair																									
	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**
Lettre de la clé	Lettres chiffrées (au croisement de la colonne *Lettre en clair* et de la ligne *Lettre de la clé*)																									
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

Figure 2 : table de Trithème utilisée dans la méthode de Vigenère

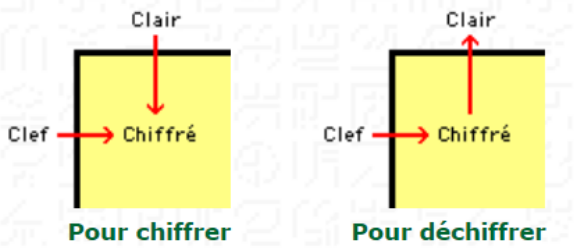

Figure 3 : méthode du chiffre de Vigenère classique

Cette méthode d'encryptage constitue la plus simple des variantes. Le message chiffré est obtenu en combinant chaque lettre du message clair avec chaque lettre de la clef.
Pour chiffrer, la lettre du message clair est lue sur la première ligne, et celle de la clef est lue sur la première colonne. La rencontre de ces deux lettres sur la matrice donne le caractère chiffré. Pour déchiffrer le message, le même schéma est gardé, et la connaissance de la clef et du message chiffré permet le déchiffrement.

- Méthode du chiffre de Beaufort

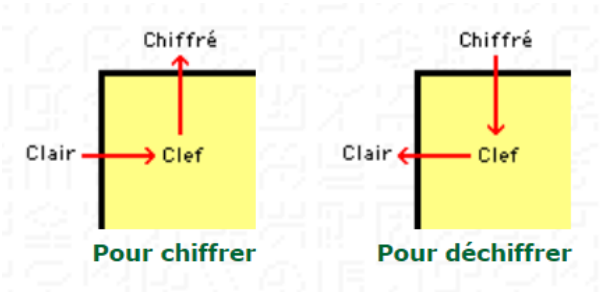

Figure 4 : méthode du chiffre de Beaufort

Dans cette variante du chiffre de Vigenère, les axes de lecture du chiffre et du message clair sont simplement inversés. Le principe reste le même. La méthode est illustrée dans la figure 4 ci-dessus.

- Variante à l'allemande du chiffre de Beaufort

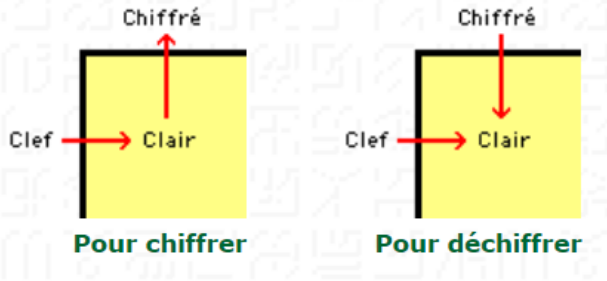

Figure 5 : Méthode de la variante à l'allemande du chiffre de Beaufort

Cette variante résulte encore une fois d'une autre façon de lire la table de Trithème. De même, la méthode de lecture est illustrée dans la figure 5 ci-dessus.

- Variante de Rozier

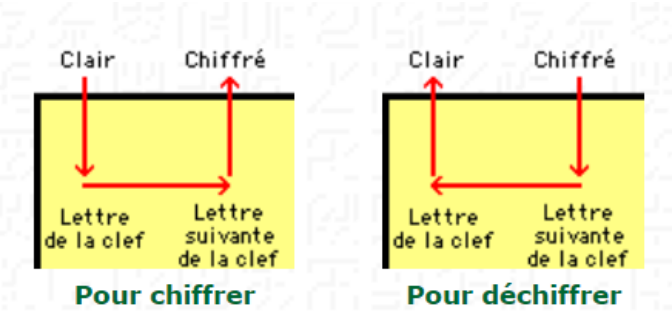

Figure 6 : Méthode de la variante de Rozier

Cette méthode de chiffrement possède un schéma de lecture significativement des autres méthodes puisque lettre du message clair et chiffré sont toutes deux lues sur la première ligne du tableau. D'autre part, la clef est utilisée deux fois : une première fois pour déterminer une lettre de la clef, et une seconde pour obtenir la lettre suivante. Bien que paraissant plus complexe, la méthode de Rozier se ramène à un chiffre de Vigenère simple en réalisant une permutation circulaire d'un cran sur la clef.

- Méthode du chiffre de Gronsfeld

Cette méthode est basée sur l'utilisation d'une clef numérique, et non plus d'une chaine de caractères. La clef est adaptée à la taille du message, répétée ou recoupée si besoin. Chaque nombre correspondra

à un décalage à imposer à chaque lettre du message clair. De cette façon, en utilisant les index de l'alphabet, il est possible chiffrer le message. Si l'index dépasse les bornes données par l'alphabet : 0 et 24, alors il faut repartir de la fin ou du début de celui-ci. Pour cela, deux façons de coder sont possibles : soit l'on créé un alphabet répété plusieurs fois (un message codé dépassera rarement six fois la longueur de l'alphabet par exemple), soit l'on créé une boucle sur celui-ci, qui renvoi le bon index. Les deux méthodes sont proposées dans le code.

- Procédé autoclave

Ce chiffrement utilise une clef déterminée grâce au message clair. Une lettre est choisie et placée en première position devant celui-ci et constitue la clef. L'utilisateur qui chiffre a donc seulement besoin de choisir une lettre, et celui qui déchiffre doit connaître le message clair et sa lettre associée. Cette méthode est donc peu utile dans le domaine de la cryptographie, mais sera tout de même codée.

- Méthode du chiffre de Porta

La méthode de substitution poly-alphabétique de Porta est basée sur l'utilisation d'une liste de douze alphabets répartis sur deux lignes de 13 caractères. En réalité, un unique alphabet est nécessaire à la construction de cette grille, puisque tous les autres dérivent du premier : le second alphabet est une permutation circulaire de la seconde ligne du premier alphabet. La grille est la suivante :

AB	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
CD	a b c d e f g h i j k l m	z n o p q r s t u v w x y
EF	a b c d e f g h i j k l m	y z n o p q r s t u v w x
GH	a b c d e f g h i j k l m	x y z n o p q r s t u v w
IJ	a b c d e f g h i j k l m	w x y z n o p q r s t u v
KL	a b c d e f g h i j k l m	v w x y z n o p q r s t u
MN	a b c d e f g h i j k l m	u v w x y z n o p q r s t
OP	a b c d e f g h i j k l m	t u v w x y z n o p q r s
QR	a b c d e f g h i j k l m	s t u v w x y z n o p q r
ST	a b c d e f g h i j k l m	r s t u v w x y z n o p q
UV	a b c d e f g h i j k l m	q r s t u v w x y z n o p
WX	a b c d e f g h i j k l m	p q r s t u v w x y z n o
YZ	a b c d e f g h i j k l m	o p q r s t u v w x y z n

Tout comme les autres méthodes poly-alphabétiques, le chiffre de Porta utilise une clef dont les lettres donnent l'alphabet utilisé. Par exemple, si la clef est « LOCK », alors la première lettre du message clair sera tirée de l'alphabet « L » (6 -ème ligne). Ensuite, il faut prendre la lettre au-dessus ou en dessous de celle du message clair dans cet alphabet. Le processus est réitéré pour chaque combinaison lettre / clef.

Pour une meilleure sécurité, il vaut mieux utiliser des alphabets composés de 26 lettres réparties aléatoirement. Pour une question de simplicité, on prendra ici l'alphabet classique.

Le déchiffrement du code est issu du même processus de chiffrage, donc il n'y a pas besoin de méthode de déchiffrement pour cette classe.

Figure 7 : Table utilisée dans la méthode du chiffre de Porta

- Méthode du chiffre de Bellaso

Belasso utilise un système légèrement moins élaboré que Porta tout en restant sur le même principe : une table contenant divers alphabets, tous constitués à partir d'un unique premier permet de chiffrer un message grâce à un clef. Cependant, Belasso travaille en termes de mots.

En effet, il code un mot à partir d'un alphabet et donc d'une lettre de la clef. Ainsi, la longueur de la clef est le nombre de mots. Il est donc nécessaire de conserver les espaces du texte d'entrée afin de les distinguer.

Notons qu'à l'époque de Bellaso, l'alphabet ne contient que 21 lettres, et celui-ci a fusionné U et V pour pouvoir fabriquer 5 alphabets. La table de Bellaso est donnée ci-dessous.

Puisque l'alphabet actuel contient 26 lettres, il est complexe de fabriquer des alphabets déterminés par quatre lettres. Une solution a été de reprendre la table de Porta en regroupant les alphabets pour quatre caractères, et deux pour les deux dernières lettres. La méthode en est modifiée mais garde le regroupement de quatre lettres pour les lettres les plus courantes.

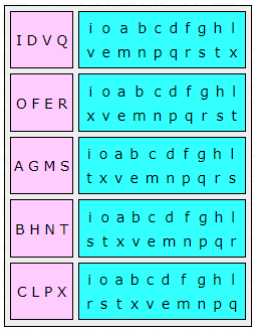

Figure 8 : table utilisée dans la méthode de Bellaso

Les chiffres polygraphiques

- Chiffre de Hill

Chaque caractère est remplacé par un nombre n (soit son rang diminué de 1 dans l'alphabet, soit son code ASCII, diminué de 32). Les caractères sont ensuite regroupés par bloc de p caractères. Les valeurs des caractères définissent alors un vecteur X. Une matrice A de taille p x p est ensuite choisie. La valeur du déterminant de la matrice est sélectionnée de façon qu'il soit premier avec la longueur de l'alphabet utilisé. Les autres termes de la matrice sont choisis de façon qu'elle soit inversible modulo n, ce qui est indispensable pour pouvoir décrypter le message. Les vecteurs Y sont ensuite calculés tels que : AX=Y

Enfin en remplaçant les valeurs dans ces vecteurs par leur lettre, un nouveau message codé est trouvé. Prenons un exemple avec p=2 et n=26 :

Valeurs : 19 ; 4 ; 23 ; 19 ; 4 ; 0 ; 2 ; 7 ; 8 ; 5 ; 5 ; 17 ; 4 ; 17

Vecteurs : X1 = (19 ; 4) ; X2= (23 ;19) ; X3= (4;0); X4=(2;7); X5=(8;5); X6=(5;17); X7=(4;17)

Vecteurs Y obtenus : (25 ; 0) ; (8 ; 19) ; (12 ; 24) ; (15 ; 1) ; (23 ; 3) ; (22 ; 7) ; (19 ; 2)

Nouveau message : ZAITMYPBXDWHTB.

Le décryptage selon la méthode du chiffre de Hill consiste à inverser la matrice A. Il est alors possible de remonter aux vecteurs X initiaux. Pour décrypter un message codé par le chiffre de Hill sans connaître la matrice à inverser, il faut alors travailler sur les fréquences de paquet.

- Chiffre de Collon

Le chiffrement selon la méthode de Collon nécessite une grille (5x5 généralement), et un nombre N donné. On sépare le message en série de N lettres. Pour chaque lettre, on cherche sa position dans la grille, pour ensuite trouver la lettre située la plus à gauche de la ligne et celle situé le plus en bas de la colonne. Ces deux nouvelles lettres remplaceront leur lettre d'origine dans le nouveau message. L'ordre appliqué pour le message codé est le suivant : les N lettres de ligne, puis les N lettres de colonne et ensuite une autre série est accolé.

Par exemple, DCODE donne DC OD E donnant AAYXLAYYAZ une fois encodé.

Pour de décodage, on applique cette méthode dans le sens inverse. Le décodage nécessite de connaître N. La connaissance de la grille utilisée n'est pas obligatoire car il existe une méthode permettant de la retrouver.

Aussi le message est séparé en 2N séries, elles-mêmes scindées en 2. Ensuite la N-ième lettre est prise pour former des paires. Il faut ensuite trouver l'intersection entre la ligne contenant la première lettre et la colonne contenant la seconde.

- Chiffre des 4 carrés

La méthode utilise 4 carrés disposé comme le montre l'illustration ci-dessous. Les carrés du haut à gauche et du bas à droite contiennent les 25 lettres de l'alphabet (j souvent exclu), dans l'ordre alphabétique. Les deux autres contiennent également ces lettres mais disposées aléatoirement.

La méthode consiste à découper le mot en paquet de deux lettres. La première est placé dans le carré ordonné du haut la seconde celui ordonné en bas. Deux droites verticales et horizontales sont ensuite tracé pour chaque lettre. L'intersection dans le carré désordonné du haut donne la première, la seconde se trouvant dans le carré désordonné du bas.

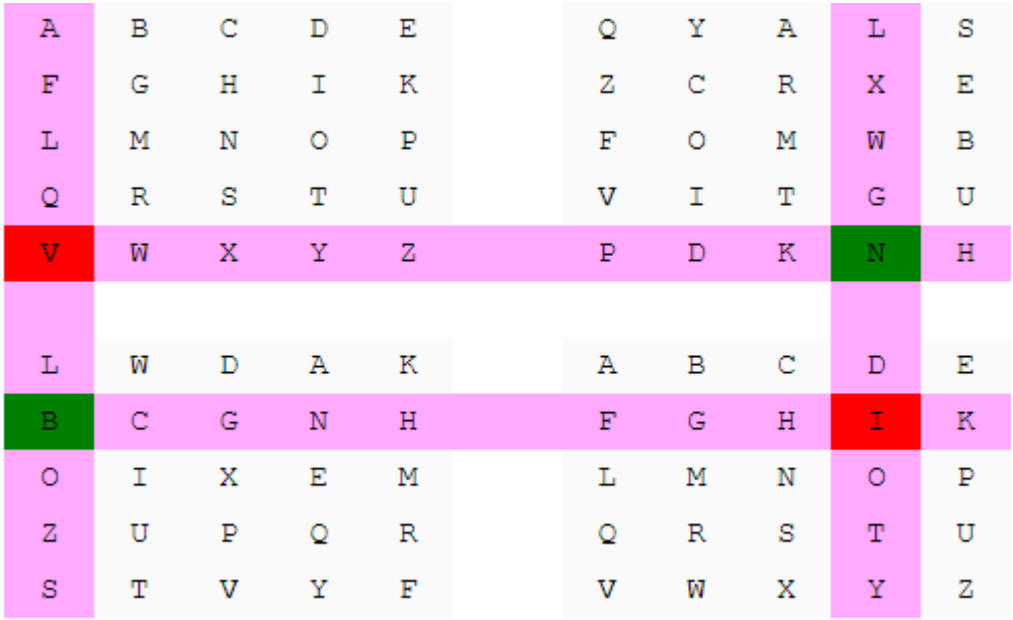

Figure 9 : Illustration du fonctionnement de la méthode des 4 carrés

Pour le décryptage d'un message par la méthode des 4 carrés, on applique strictement la méthode décrite ci-dessus dans le sens inverse.

III. Méthodes de cryptanalyse

La cryptanalyse est l'art d'analyser un message chiffré afin de le décrypter. Il convient donc de l'y intéresser dans le cadre d'un projet de cryptographie.

Un des moyens les plus simples de détecter le chiffre d'un message est l'analyse de fréquences. Elle consiste à déterminer les fréquences d'apparition des lettres dans un message, et à les comparer à celles de l'alphabet d'une langue donnée. En effet, les fréquences des lettres varient d'une langue à une autre. Pour ce faire, le message devant être comparé à l'alphabet doit être suffisamment long pour que les moyennes soient significatives. Le texte de référence a été pris sur le site projet

Gutemberg, les livres sont donc libres d'accès. Une fonction s'occupe de traiter le message pour supprimer les caractères spéciaux, et les lettres sont ensuite comptées. Cette analyse se restreint cependant à des chiffres simples comme les méthodes monoalphabétiques. Elle n'est plus applicable aux chiffres monoalphabétiques comme la méthode de Vigenère, qui lisse les fréquences. Il s'agit d'un calcul lourd qui peu prendre du temps selon la longueur des textes à comparer. Une approche similaire consiste à analyser les fréquences des bigrammes du texte, en comparaison à une langue. Il faut déterminer la longueur des séquences, et décrypter par la suite à l'aide du mot probable. Les seuls histogrammes des fréquences sont abordés dans le programme.

IV. Stratégies et méthodes utilisées dans le code

Point de vue global sur le projet

L'ensemble du projet est codé en langage python, plus précisément en programmation orientée objet. Chaque méthode décrite dans la partie précédente représente une classe. Chacune de ces classes présentent plusieurs méthodes dont au moins une appelée « chiffrement » et une autre appelée « déchiffrement ». Chaque classe renvoie un résultat appelé simplement « result », il s'agit du mot codé ou du mot décodé selon le choix de l'utilisateur.

Le choix de l'utilisateur est au cœur de ce projet de cryptographie. En effet, une interface graphique a été codée grâce au module PyQt5 de python. Par l'intermédiaire de multiples méthodes codées au sein de l'interface graphique, la liaison entre l'interface et les classes des méthodes mono, poly-alphabétiques et polygraphiques a été faite.

Choix et stratégies spécifiques à l'interface graphique

Le but est de proposer à l'utilisateur une interface simple, facile d'utilisation et aussi complète que possible.

Pour cela nous avons choisi de présenter les différentes méthodes sous la forme d'un menu déroulant ce qui permet non seulement de sélectionner la méthode souhaitée, d'éviter l'affichage de trop d'informations sur l'interface mais également d'assurer un choix par défaut si l'utilisateur oublie d'en sélectionner un.

De plus, des messages d'informations de style « pop-up » ont été mis en place dans le but d'informer l'utilisateur des conditions que doivent respecter sa clé et/ou son entrée en fonction de la méthode de codage ou de décodage qu'il a sélectionné.

Des boutons à choix unique permettent à l'utilisateur dans quel sens il souhaite utiliser le programme. Le « choix unique » est important, nous avions dans un premier temps choisi des « checkbuttons » mais ces derniers permettent des choix multiples qui auraient fait dysfonctionner notre programme.

Enfin, des boutons « information » sont disponibles, ils permettent à l'utilisateur d'obtenir des informations sur la méthode qu'il est sur le point d'utiliser en passant sa souris sur le bouton.

Problèmes rencontrés

Lors de la programmation des classes

Chaque classe est contenue dans un fichier au format py. Une classe contient une unique méthode de cryptographie. Ce choix possède l'avantage d'être extrêmement clair en termes de code (import, nomenclature …), mais a l'inconvénient de multiplier les fichiers. Ainsi, nous avons une vingtaine de fichiers dans le dossier pycharm. Malgré ce problème, nous avons remarqué qu'il était facile de

généraliser les cas dans le GUI grâce à cette standardisation des classes. Certaines méthodes se prêtaient cependant moins à l'utilisation des classes.

Lors de la programmation de l'interface graphique

L'interface graphique est assez redondante en termes de programmation, même si l'objectif est différent à chaque fois. Nous aurions pu utiliser QtDesigner mais nous avons préféré contrôler les effets de chacune de nos actions sur les placements de chaque objet.

Le code de l'interface est long (environ 550 lignes). Il était donc difficile de s'y retrouver. Nous avons dû utiliser l'affichage compacte de pycharm sur certaines fonctions peu éditées. À l'avenir, il serait donc préférable de coder plusieurs classes dans différents fichiers.

Lors de la mise en relation GUI – class

La méthode de codage est propre à chacun, ce qui provoque des variantes parfois peu adaptables, par exemple pour le code de Collon et des quatre carrés. La liaison avec le GUI a donc été plus compliquée que prévue. Dans certains cas, elle n'a pas été faite du fait du manque de temps.

Les méthodes mono-alphabétiques

- ⇨ Méthode de Atbash et de César

Ces deux méthodes sont les plus basiques de toutes. Elles n'ont pas nécessité de fonctions ou de bibliothèques spécifiques pour les coder, simplement l'utilisation d'une boucle « for » comprenant une ou plusieurs boucles « if », l'ensemble permettant de former le mot souhaité lettre par lettre en ajoutant les caractères à une liste initialement vide. A la fin de chaque méthode, l'utilisation d'une fonction particulière permet de présenter le résultat sous forme de mot (chaine de caractères) et non de liste. L'ensemble des méthodes de Atbash et de César codées sont capables de prendre en compte les espaces.

- ⇨ Méthode du carré de Polybe

Cette méthode est plus complexe, Les codes pour encrypter ou décrypter sont très différents. Le premier étant similaire à ceux des deux méthodes ci-dessus (ajout caractère par caractère à une liste initialement vide), alors que le deuxième est beaucoup plus court car il ne consiste qu'en une lecture de coordonnées (en une seule ligne) et à un affichage du caractère associé. Notez que l'encryptage renvoie un code constitué de chiffres (deux fois plus que les lettres du mot de base car on a deux coordonnées par lettre) et différencie les I des J alors que le décryptage renvoie un mot et ne différencie pas les I et les J. L'ensemble de la méthode de Polybe codée prend en compte la présence d'espace.

- ⇨ Code du Che

La méthode codée sous le nom de code du Che est en fait une variante du code original. En effet, certains caractères étant codés par des nombres de 1 chiffres et d'autres par des nombres de 2 chiffres, le fait de découper le message codé en blocs de 5 chiffres peut avoir pour effet de couper en deux le code d'une lettre. Pour pallier cet effet, le programme cryptant et décryptant le code bloc par bloc, il a été décidé de créer des blocs de 6 chiffres si une telle situation se présente. Le caractère « en trop » est alors additionné avec la valeur 0.

Concernant le code des espaces, il a été décidé de leur attribuer le code 55. Puisque ce code et le code 5 ne sont pas utilisés pour coder d'autres lettres, ils ne risquent pas d'être confondus lors du décryptage.

De même, les codes 3 et 33 n'étant pas utilisés, il a été décidé de remplir avec des 3 le dernier bloc si celui-ci n'est pas rempli par 5 chiffres. Cela permet alors de bien réaliser l'addition avec la clé à 5 chiffres modulo 10, et le décryptage n'est pas altéré puisqu'il est impossible de confondre ces codes avec d'autres.

Les méthodes poly-alphabétiques

- ⇨ Méthodes de Vigenère et dérivées

Ces méthodes sont basées sur la manipulation d'indexes de listes et de chaines de caractères. Il s'agissait de transcrire la lecture de la table de Trithème (Figure 2) en un code python : la manipulation d'indices sur des alphabets est donc le point clé ici.

- ⇨ Autoclave

Le procédé autoclave n'étant rien d'autre qu'un code de Vigenère avec un clef prédéfinie, rien ne sera développé dessus. La méthode est fonctionnelle dans le code mais sans intérêt.

- ⇨ Porta et Bellaso

Ces deux dernières méthodes des substitutions polyalphabétiques sont basées sur l'utilisation d'une table. Il a donc été nécessaire de la modéliser. L'utilisateur aurait dû la saisir, puisqu'il s'agit d'un élément déterminant dans la fiabilité du code, mais par souci de complexité, celle-ci a été prise par défaut afin de réaliser le code. Dans une version plus aboutie, nous aurions proposé à l'utilisateur de rentrer les alphabets de son choix pour chaque lettre ou bien d'accepter d'utiliser une table générée aléatoirement. Une propriété commune à ces deux méthodes est qu'une seule méthode chiffrage est nécessaire. En effet, un chiffrage sur un code chiffré par Porta ou Bellaso le décodera. Cela a radicalement simplifié le code.

Les méthodes polygraphiques

- ⇨ Chiffre des quatre carrés

Cette méthode nécessite un partitionnement du message, aussi bien pour le codage que pour le décodage. Une fonction permettant la division en binôme auxquels sera appliquée la méthode des chiffres des quatre carrés à proprement parler, est codée. Celle-ci permet également d'ajouter un « A » par défaut si le mot possède un nombre de lettres impaire. Une fonction d'analyse permet d'indiquer si c'est le cas. Une méthode de génération de grille désordonnée est également créée et ceux en dehors de la classe. En effet cette méthode aurait pu être réutilisée autre part.

- ⇨ Chiffre de Collon

La méthode du chiffre de Collon a nécessité deux fonctions de partitionnement, dues au décodage légèrement différent du codage. Une fonction en dehors de la classe permet de trouver les coordonnées d'une lettre dans la matrice des lettres ordonnées.

Remarque sur les deux dernières méthodes :

Le « J » est toujours remplacé par un « I » ce qui peut amener à quelques légères erreurs .

- ⇨ Chiffre de Hill

Cette méthode nécessite de couper le message en blocs de p lettres, qui seront ensuite codés à l'aide d'une matrice de taille p x p. Or, cette matrice doit être inversible, et ses coefficients doivent être entiers (de la matrice et de la matrice inverse), puisqu'il est difficile de trouver la 1/3ème lettre de l'alphabet. Ces conditions étant difficile à obtenir, il n'a été codé que la méthode prenant p = 2. En

effet, ce cas rend plus simple les conditions d'inversion de la matrice et donc les conditions sur les paramètres choisis par l'utilisateur (et donc la vérification de celles-ci).

Si la longueur du message ne permet pas de remplir le dernier bloc en entier, celui-ci est alors complété par des zéros. Dans le cas p = 2, le code de la dernière lettre sera alors la a * x (mod(len(alphabet))) ème lettre de l'alphabet. Il est alors possible de traiter ce cas séparément pour remonter à la lettre d'origine lors du décryptage.

Choix concernant l'interface graphique

L'application étant globalement fonctionnelle, nous avons choisi de la rentre visuellement correcte.

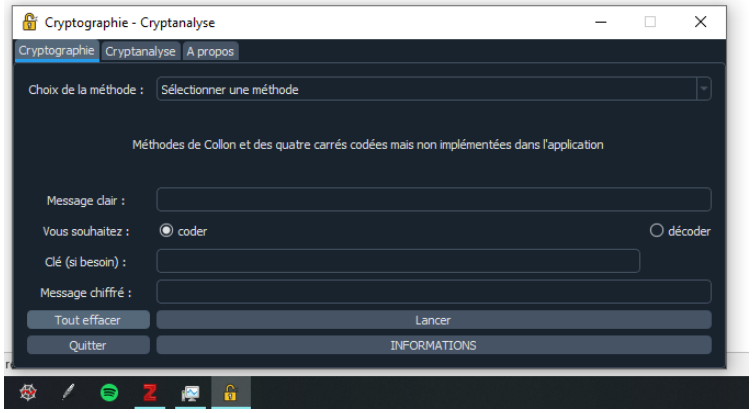

Figure 10 : Aperçu de l'application montrant le thème sombre, et les icônes utilisées

Pour cela, le module qdarkstyle a été utilisé. Il s'agit d'un thème sombre pour une application de PyQt5. D'autre part, une icône d'application a été choisie pour l'affichage dans la barre des tâches ainsi qu'au niveau de la fenêtre (Figure 10).

Afin d'éviter les effets d'échelle, la dimension de la fenêtre a été fixée.

V. Notice d'utilisation de l'application

Prérequis

Avant de lancer l'application, s'assurer que les modules :

- ✓ qdarkstyle
- ✓ matplotlib
- ✓ qtpy

sont installés. Dans le cas contraire, le programme ne fonctionnera pas.

Onglet de cryptographie

L'application est composée de trois menus. Le menu par défaut apparaissant à son ouverture est celui de la cryptographie (Figure 11).

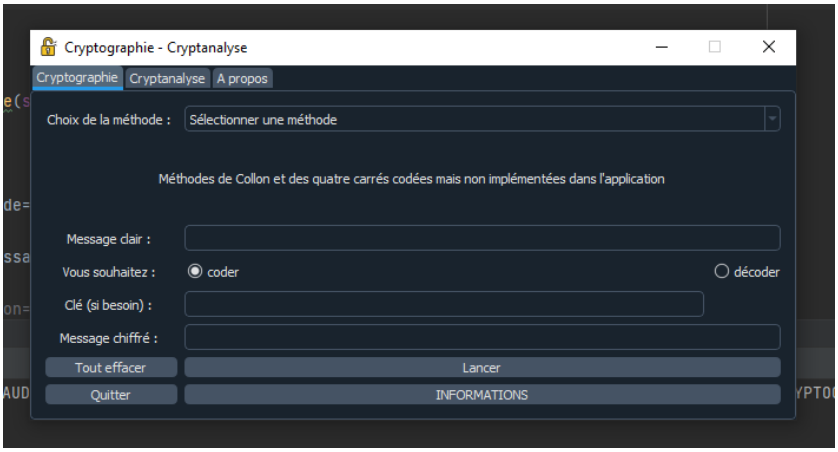

Figure 11 : Organisation de l'onglet cryptographie

En première ligne, l'utilisateur peut choisir une méthode de chiffrement parmi celles qui sont disponibles. Certains choix ne sont que des séparateurs et mèneront donc à un message d'information.

En dessous, un label communique une information globale sur l'état de certaines méthodes : ici, deux méthodes n'ont pas encore été implémentées. Lorsqu'une méthode est sélectionnée, ce label affiche des informations sur la saisie.

Par la suite, l'utilisateur doit entrer un message clair pour le chiffrer. Certaines méthodes ne nécessitent pas de clé ou alors demandent une clé d'un certain type. Il convient de remplir à minima deux de ces champs. Dans le cas contraire, un message d'erreur s'affiche.

Par défaut, le codage est sélectionné, si l'on veut décoder, il suffit de cocher décoder. Il faut donc alors remplir la clé (si besoin) et le message chiffré, tout en sélectionnant la méthode souhaitée.

Le bouton lancer lance le processus de chiffrement si les champs sont correctement remplis.

Le bouton tout effacer efface toutes les zones de saisie des textes de l'application simultanément.

Le bouton quitter ferme l'application.

Le bouton information permet d'obtenir des données générales sur la méthode utilisée.

Onglet cryptanalyse

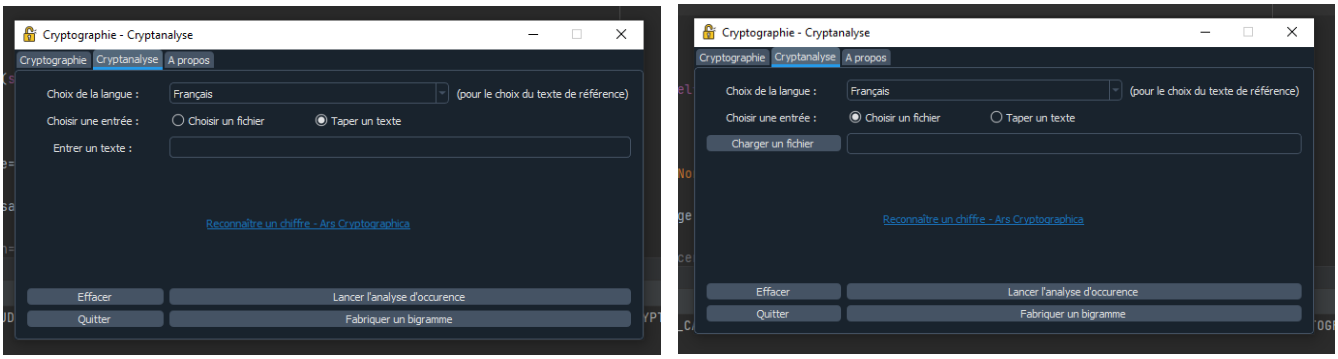

Figure 12 : Organisation de l'onglet cryptanalyse

L'onglet cryptanalyse (Figure 12) est un outil permettant de générer des histogrammes de fréquence des lettres dans un message.

Dans un premier temps, l'utilisateur est invité à choisir une langue de référence (entre français et anglais). Ce texte est issu du site projet Gutenberg. Il s'agit d'un livre écrit dans la langue choisie. L'échantillon est supposé assez grand pour que l'analyse soit généralisée à une langue. Cela correspondra au texte de référence dans l'histogramme (Figure 13).

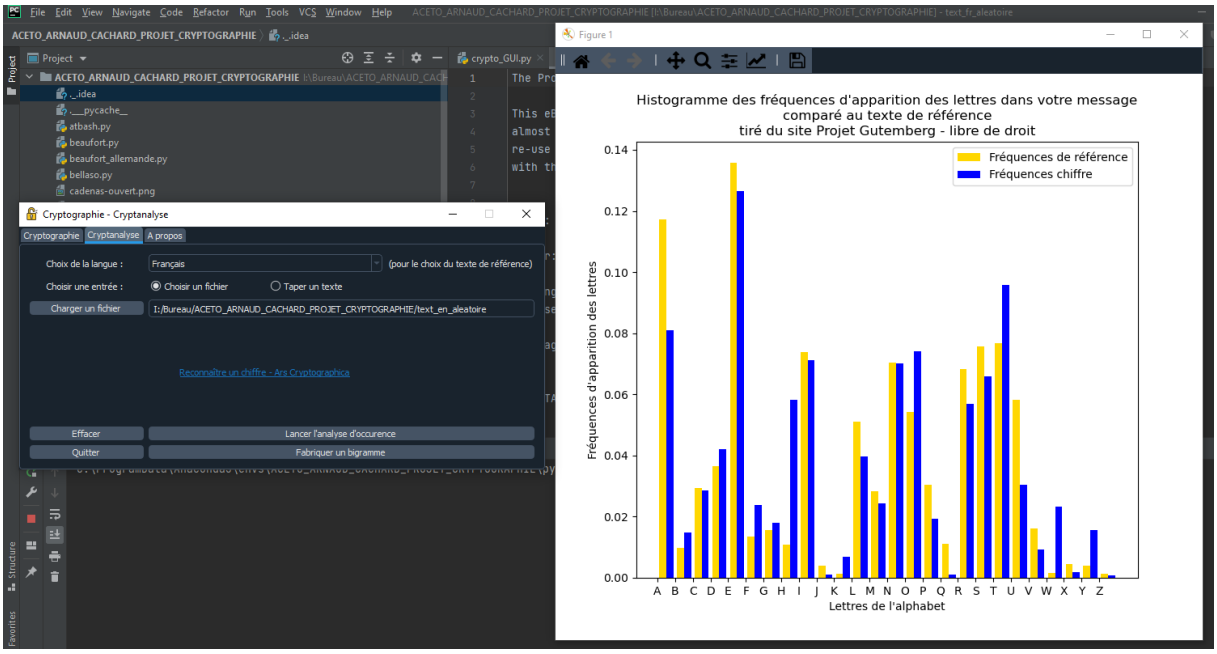

Figure 13 : Fonction histogramme de l'onglet cryptanalyse

Un type de saisie doit ensuite être choisi (par défaut, il s'agit d'une entrée manuelle).

L'utilisateur peut donc taper un texte chiffré dont il ne connaît pas la méthode de chiffrement, puis lancer l'analyse d'occurrence, qui affichera un histogramme.

Le bouton « Fabriquer un bigramme » n'est pas encore fonctionnel.

Une autre façon d'utiliser cet onglet est d'aller chercher un fichier texte (.txt) à l'aide du menu. L'analyse compare alors les fréquences du texte de référence au fichier lu.

Quitter et effacer ont les mêmes effets que précédemment.

Onglet d’informations sur le programme

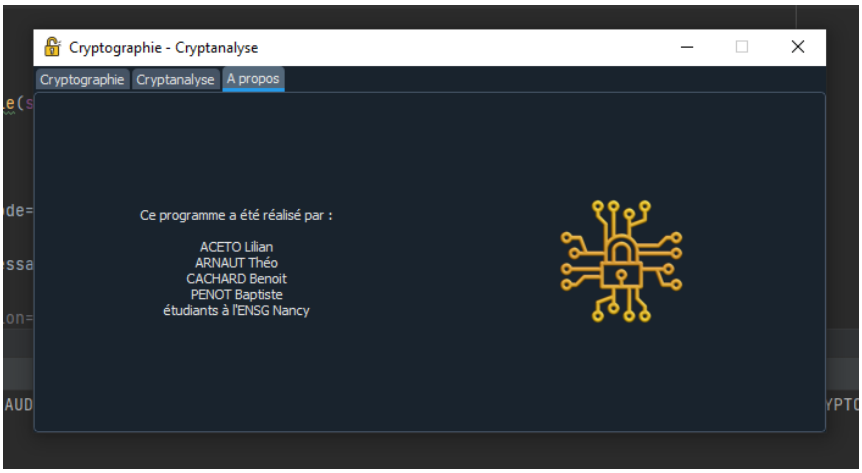

Figure 14 : Organisation de l'onglet "à propos"

Le dernier onglet (Figure 14) est vocation esthétique et informative.

VI. Bibliographie et documentation utilisée

- La plupart des illustrations et tableaux de la partie : La substitution poly-alphabétique provient du site Ars Cryptographica (URL : https://www.apprendre-en-ligne.net/crypto/index.html) : Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 8.

Ars Cryptographica: Table des matières [WWW Document], n.d. URL https://www.apprendre-en-ligne.net/crypto/index-crypto-subst.html

Ars Cryptographica: Table des matières [WWW Document], n.d. URL https://www.apprendre-en-ligne.net/crypto/index-crypto-transpo.html

Ars Cryptographica: Table des matières [WWW Document], n.d. URL https://www.apprendre-en-ligne.net/crypto/index-crypto-moderne.html

Cryptographie et cryptanalyse : Ars Cryptographica [WWW Document], n.d. URL https://www.apprendre-en-ligne.net/crypto/

Le tableau de Trithème [WWW Document], n.d. URL https://www.apprendre-en-ligne.net/crypto/tritheme/polyalpha.html

- Les informations utilisées proviennent majoritairement de ce même site, bien qu'elles aient été complétées par les pages Wikipédia de ces méthodes de chiffrement. Etant aisément retrouvables, elles ne seront pas listées ici.

Wikipédia:Accueil principal, 2019. . Wikipedia, the free encyclopedia.

- Les fréquences de références des lettres proviennent de l'analyse de textes libres d'accès.

Projet Gutenberg, 2021. . Wikipédia.
Project Gutenberg [WWW Document], n.d. . Project Gutenberg. URL https://www.gutenberg.org/

Verne, J., 1997. Le tour du monde en quatre-vingts jours.
Verne, J., 1994. Around the World in Eighty Days.

- Ce site a également été utilisé pour des illustrations et définitions : DCODE, https://www.dcode.fr (consulté le 23 novembre 2021)

- Les différentes commandes ont été trouvées sur la documentation tutorialspoint de PyQt5.

pyqt5_tutorial.pdf, n.d. URL https://www.tutorialspoint.com/pyqt5/pyqt5_tutorial.pdf

- Le reste provient du cours enseigné à l'ENSG Nancy sur les interfaces graphiques.