Cryptographie

Chiffre de César

On dit que César cryptait ses messages afin de les rendre incompréhensibles si quelqu'un venait à les intercepter. Sa méthode était simple : il décalait les lettres de trois rangs. Ainsi, la lettre A était associée à un D, la lettre B au E et ainsi de suite. Bien sûr, pour crypter les trois dernières lettres, il suffisait de recommencer l'alphabet.

1. On note x le rang de la lettre à crypter (on propose d'associer la lettre A au nombre 0 pour plus de simplicité) et y le rang de la lettre chiffrée. Proposer une formulation afin d'exprimer la relation entre x et y.
2. Le code de César est un mode de chiffrement affine. A chaque lettre est associée une autre selon une relation mathématique de type $y = ax + b$, x et y étant les rangs associés à ces lettres (ou plus précisément, le rang de la lettre après cryptage correspond au reste de y dans la division euclidienne par 26).
   On considère dans un premier temps que $b = 0$.
   (a) On propose d'utiliser $a = 27$. Qu'en pensez-vous ?
   On pourra utiliser une fonction de la calculatrice pour calculer plus rapidement le reste d'une division euclidienne : Boîte à outils > Arithmétique.
   (b) Quelle condition faut-il respecter pour que x et y ne soient pas associés au même rang, c'est-à-dire qu'une lettre ne puisse pas être codée par elle-même ?
3. On considère maintenant que b est un réel non-nul.
   (a) On propose de crypter un message à l'aide de la relation $y = 2x + 1$. Que constate-t-on ?
   (b) On veut que chaque lettre soit codée de manière différente. C'est-à-dire que pour chaque lettre de l'alphabet de l'ensemble de départ, on veut associer une lettre de l'alphabet dans l'ensemble d'arrivée de sorte à ce qu'elles soient toutes utilisées. Chaque lettre codée n'aura pour antécédent qu'une seule autre lettre. C'est ce que l'on appelle une **bijection**.
   Pour respecter cette condition, il faut que le coefficient a soit premier avec le nombre de lettres utilisées. Proposer un type de chiffrement affine qui vérifie cette condition.
4. On propose de crypter le mot MATH en utilisant une fonction f définie telle que pour tout x compris entre 0 et 25, correspondant au rang de la lettre de départ, on associe $y = 7x + 12$. Le reste de la division euclidienne de y par 26 donne le rang de la lettre chiffrée. Comment est crypté le mot MATH ?

Pour aller plus loin !

Pour augmenter d'un cran la sécurité de notre cryptage, il est possible de procéder à un cryptage en utilisant une clé, sous forme d'un mot. C'est la clé qui va déterminer la substitution nécessaire. Prenons un exemple.

On souhaite toujours crypter le mot MATH mais cette fois-ci, nous allons utiliser une clé : CHAT. Le cryptage de la lettre M se fait à l'aide de la lettre C, de rang 2, ce qui détermine le nombre de décalage à effectuer : la lettre M devient O. De la même façon, la lettre A est cryptée par un H, de rang 7 : elle devient elle-même un H. La lettre T est cryptée par la lettre A, de rang 0 : il n'y a pas de décalage, le T reste un T. Enfin, la lettre H est cryptée par un T, de rang 19 : elle devient un A.
Le mot MATH se code OHTA.
Ce type de chiffrement est appelé chiffre de Vigenère, du nom de son génial inventeur. Pour coder ou décoder plus rapidement un message en suivant ce principe, on peut se référer à la table de Vigenère, qui se trouve facilement sur Internet.

Décrypter par analyse fréquentielle

Dans la langue française, certaines lettres sont beaucoup plus courantes que d'autres. Par exemple, la lettre "E" est de loin l'une des plus utilisées.
Il est possible de calculer approximativement la fréquence d'apparition de chacune des lettres dans notre langue à partir d'une grande quantité de texte. Cette fréquence peut ensuite être utilisée pour déchiffrer un texte chiffré par substitution.
Wikipédia nous donne le tableau suivant (les valeurs sont données en pourcentage) pour la langue française :

A	B	C	D	E	F	G	H	I	J	K	L	M
8,20	0,90	3,35	3,67	16,72	1,07	0,87	0,74	7,58	0,61	0,07	5,46	2,97
M	O	P	Q	R	S	T	U	V	W	X	Y	Z
7,10	5,84	2,52	1,36	6,69	7,95	7,24	6,43	1,84	0,05	0,43	0,13	0,33

1. Combien de fois, en moyenne, apparaît la lettre E, dans une page de 2500 caractères ?
2. On souhaite décrypter un texte codé. Calculer la fréquence d'apparition de chacune des lettres dans ce texte codé puis décoder le texte en proposant des pistes d'amélioration de la méthode par analyse fréquentielle.
   La machine est quelquefois plus fiable que l'humain ! Aussi pour calculer les fréquences, on propose de programmer un script en Python. On pourra notamment utiliser l'instruction `count()` qui permet de compter le nombre d'occurence dans une chaîne de caractères. Par exemple, si 'message' contient une chaîne de caractères, l'instruction `message.count(ma)` comptera le nombre d'occurences de "ma" dans cette chaîne.
   Voici le texte à décoder :
   LOI SMPJOSMPQUWOI IGZP WZ OZIOSTLO HO AGZZMQIIMZAOI MTIPBMQPOI BOIWLPMZP HO BMQIGZZOSOZPI LGCQUWOI MNNLQUWOI M HOI GTXOPI HQDOBI OLI UWO LOI OZIOSTLOI SMPJOSMPQUWOI, LOI ZGSTBOI, LOI VGBSOI… SOBAQ KQEQNOHQM !