Laboratoire n°1 - CRY - Alexis Martins

Q1. Quel est l'avantage d'utiliser le test du x2 plutôt que de comparer simplement la lettre la plus fréquente dans le texte chiffré par rapport aux statistiques du langage de base ?

L'avantage de ce test est qu'il nous donne une vision globale sur l'ensemble du texte. Si on se concentrait que sur une lettre, on pourrait avoir des cas où ça ne marche pas juste parce que la lettre choisie n'est pas représentative de la langue choisie ou que le texte n'a pas la même fréquence d'apparition de cette lettre que notre texte de référence. Ainsi avec le X^2, on vérifie bien que l'ensemble du texte soit cohérent à nos attentes.

Q2. Pourquoi est-ce que le test du x2 ne fonctionne-t-il pas directement sur un texte chiffré à l'aide du chiffre de Vigenère ?

C'est une méthode de chiffrement poly-alphabétique, cela veut dire qu'une lettre en clair dans le texte n'aura pas toujours la même représentation une fois chiffrée. Cela rend donc inefficace l'analyse pour savoir la fréquence d'apparition des lettres. C'était l'avantage avec César, c'est qu'une même lettre avait toujours la même version chiffrée ainsi l'analyse de fréquence était juste décalée du décalage de la clé. Dans le cas de Vigenère le décalage n'est pas toujours le même et les fréquences ne sont donc pas une bonne indication.

> "Le chiffre de Vigenère est un système de chiffrement par substitution polyalphabétique dans lequel une même lettre du message clair peut, suivant sa position dans celui-ci, être remplacée par des lettres différentes, contrairement à un système de chiffrement mono alphabétique comme le chiffre de César (qu'il utilise cependant comme composant). Cette méthode résiste ainsi à l'analyse de fréquences, ce qui est un avantage décisif sur les chiffrements mono alphabétiques." - Wikipedia

Q3. Que mesure l'indice de coïncidence ?

L'indice de coïncidence représente la probabilité d'avoir deux lettres identiques lors d'un tirage aléatoire dans le texte. Plus celui-ci est bas plus le texte est aléatoire. C'est aussi grâce à cet indice que l'on peut déterminer la longueur de la clé utilisée pour chiffrer le texte.

Q4. Pourquoi est-ce que l'indice de coïncidence n'est-il pas modifié lorsque l'on applique le chiffre de César généralisé sur un texte ?

César est un chiffrement mono-alphabétique, donc une lettre en clair aura toujours la même représentation une fois chiffrée. Ainsi les probabilités restent les mêmes, mais décalées selon la clé utilisée. La seule différence, c'est que ces probabilités sont maintenant décalées d'un décalage de clé. Cela n'a donc aucun impact sur l'indice de coïncidence.

Q5. Est-il possible de coder un logiciel permettant de décrypter un document chiffré avec le chiffre de Vigenère et une clef ayant la même taille que le texte clair ? Justifiez.

Non, ce n'est pas possible car cela voudrait dire que chaque lettre a été chiffré avec une lettre différente. On ne retrouverait plus le pattern cyclique que l'on faisait avec Vigenère qui était de prendre toutes les lettres espacées de la taille de la clé pour avoir un chiffre de César à casser.

Ce qui implique qu'il n'y a plus de fréquence d'apparition des lettres qui se rapprocherait du français et qu'il n'est pas possible non plus d'essayer de deviner la taille de la clé en prenant des lettres espacées de la longueur de la clé pour essayer de retrouver un texte avec un indice de coïncidence proche de celui du français.

On se rapproche alors énormément du masque jetable vu en cours qui est inconditionnellement sûr. La seule différence étant que pour un masque jetable, il faut une clé utilisée une seule fois et totalement aléatoire ce que la question ne précise pas. Mais dans ce cas-là, c'est en effet impossible de retrouver la clé.

Q6. Expliquez votre attaque sur la version améliorée du chiffre de Vigenère.

La première partie consiste à trouver la longueur de la clé, ainsi que la clé de César utilisée. Pour cela j'ai donc fait toutes les possibilités entre la taille de la clé et le décalage de César. Pour chaque possibilité, le but était donc de déchiffrer des blocs de la taille de la clé avec le décalage de César. Plus on avance dans le texte plus de fois il faut déchiffrer le même bloc. Voici un petit exemple :

Texte chiffré : "ABCDEF"

Imaginons que nous sommes en train de tester une clé de taille 2 avec un décalage de 1. On va donc déchiffrer les blocs "AB" et "CD" avec un décalage de 1. On applique alors un décalage de 1 au premier bloc de 2 lettres, puis de 2 au second, etc...

```
Texte chiffré   : ABCDEF
Décalages       : 112233
------------------------
Texte résultant : BCEGHI
```

Le texte résultant n'est plus que théoriquement un texte de Vigenère. On y applique alors la même approche, en calculant l'indice de coïncidence pour vérifier quel texte s'approche le plus de notre texte de référence français. Il faut uniquement garder en mémoire le texte ayant l'indice de coïncidence le plus proche du français. Je prends soin de bien sauvegarder la clé de César utilisée, ainsi que la longueur de clé testée.

Une fois qu'on possède le texte de Vigenère, la clé de César et la longueur de la clé de Vigenère. Il ne reste plus qu'à appliquer un caesar_break sur chaque lettre de la clé pour la retrouver. Ainsi, on possède toutes les informations pour pouvoir ensuite déchiffrer le texte.

Q7. D'où proviennent les textes clairs correspondants aux fichiers vigenere.txt et vigenereAmeliore.txt ?

Le texte chiffré par Vigenère correspond à une description du Crunch qui est une semaine spéciale à l'HEIG ou les élèves ont le **plaisir** de travailler en groupe sur des projets du monde réel.

Le texte chiffré par Vigenère amélioré correspond à la tirade/monologue de Otis (Édouard Baer) dans le film "Astérix et Obélix : Mission Cléopâtre".