TP5 : Casser le chiffrage de Vigenère - Partie 2

1 Le TP4 : dernière chance pour le corriger et le compléter (1h30 maximum)

Vous trouverez, sur Moodle, le résultat des évaluations automatiques du TP4, sur le décodage de Vigenère. Si votre programme a été noté comme ayant des problèmes, vous pouvez le corriger dès à présent, et effectuer un nouveau dépôt de votre travail (ce sera le dernier concernant le TP4). N'oubliez pas de suivre à la lettre les instructions de rendu données dans le TP4. Vous devrez effectuer votre rendu avant 10h00.

2 Déduire la longueur du mot de passe dans un texte crypté selon Vigenère

Pour trouver la longueur du mot de passe d'un texte crypté en Vigenère, nous allons utiliser un indice statistique précieux en linguistique. L'indice de coïncidence d'un texte est la probabilité, si on prend deux lettres au hasard dans un texte, que ces deux lettres soient les mêmes.

Posons N le nombre de lettres du texte que l'on examine, N_a le nombre de 'a' dans le texte, N_b le nombre de 'b' dans le texte, etc... L'indice de coïncidence du texte est égal à

$$\frac{N_a * (N_a - 1) + N_b * (N_b - 1) + ... + N_z * (N_z - 1)}{N * (N - 1)}$$

On remarque deux choses :

— Le fait de décaler toutes les lettres d'un texte avec un même décalage, comme pour la méthode de César, ne change pas l'indice de coïncidence du texte.

— Le fait de décaler toutes les lettres d'un texte avec décalage différent, comme pour la méthode de Vigenère, change l'indice de coïncidence du texte.

Si l'on met des lettres au hasard dans un texte (le texte n'aura alors aucun sens), l'indice de coïncidence de ce dernier sera à peu près égal à 0,0385. Par-contre, si vous écrivez un texte en français (vous ne choisissez donc pas les lettres au hasard), l'indice de coïncidence sera à peu près égal à 0,0785.

Reprenons l'exemple d'un message chiffré en Vigenère avec le mot de passe "java". Si vous calculez l'indice de coïncidence de ce message, vous obtiendrez probablement un score aux alentours de 0,04 (car le message crypté ne ressemble en rien à du français). Si vous ne prenez qu'une lettre sur deux du message, l'indice de coïncidence du nouveau texte sera aussi assez proche de 0,04. Si vous ne prenez qu'une lettre sur trois du message, l'indice de coïncidence du nouveau texte sera aussi assez proche de 0,04. Par contre, si vous ne prenez qu'une lettre sur quatre du message (quatre étant la longueur du mot de passe), vous créez un message où toutes les lettres proviennent d'un texte en français et ont été décalées d'un même cran : vous constaterez alors que l'indice de coïncidence de ce nouveau texte sera très proche de 0,0785.

Pour résumer, on trouve la longueur du mot de passe utilisé pour chiffrer un texte en Vigenère en ne prenant qu'une lettre sur n dans le message, avec n variant de 1 à T (T est une borne supérieure de la longueur du mot de passe... typiquement, 30 devrait suffire). A chaque fois, calculez l'indice de coïncidence de ce sous-message, et dès que ce dernier est "proche" de 0,0785, vous avez trouvé la longueur du mot de passe : vous pouvez alors utiliser la méthode décrite dans la section précédente pour décrypter complètement le message.

Cette méthode ne fonctionne que si le texte est assez long par rapport au mot de passe utilisé. De plus, le français possédant naturellement un indice de coïncidence très élevé, cette méthode s'applique bien pour des textes en français.

Questions

1. Ecrivez une fonction **double IndiceCoincidence(FILE *in)**, qui renvoie l'indice de coïncidence du texte contenu dans le fichier **in**. On peut, pour simplifier les choses, considérer que les fichiers de texte ont été normalisés avec la fonction **normaliserTexte**.

2. Ecrivez une fonction **int longueurMotPasseVigenere(FILE *in)** qui renvoie la longueur du mot de passe ayant permis de crypter le fichier **in**. Il serait très utile d'utiliser ici la fonction **decouperFichier** précédemment écrite.

3. Proposez un programme qui décrypte automatiquement un texte passé en paramètre, et écrit son résultat dans un fichier de sortie. Est-ce que votre programme parvient bien à décrypter le texte du dernier TP ? Parvenez-vous à décrypter le texte de ce TP ?

3 Rendu de votre programme

Vous devez rendre le travail réalisé pour le TP5 sur Moodle.
Ce rendu doit suivre un format spécifique. Votre programme devra s'exécuter à l'aide de la commande

```
./vigenere_decodage mon_message_code.txt sortie.txt
```

où *mon_ message_ code.txt* est le message à décoder, et *sortie.txt* contient le message décodé par votre programme. De plus, votre programme devra compiler à l'aide de la commande

```
gcc vigenere_decodage.c -o vigenere_decodage
```

Si votre ligne de compilation est plus complexe, vous devrez alors la spécifier dans un *Makefile*. Tous vos fichiers devront être directement placés dans un fichier zip (et ne pas être dans un sous dossier du fichier zip), dont le nom sera

```
VotreNom_VotrePrenom_VotreNumeroEtudiant.zip
```

sans aucun espace (si votre nom ou votre prénom contiennent un espace, ne les faîtes pas figurer). Les formats de compression acceptés sont .zip, .tgz, .tar.gz, .7z, .rar.

Si vous avez un binôme, il vous faudra écrire son nom, prénom et numéro d'étudiant séparés par des espaces, en plus des vôtres, dans un fichier *binome.txt*, dont le format sera

```
Nom1 Prenom1 Numero_etudiant1
Nom2 Prenom2 Numero_etudiant2
```

Vous ne devez rendre le fichier qu'une seule fois par binôme.