P11 : Projet

Equipe P11

1er octobre 2021

L'objectif de ce mini-projet est de simuler le décryptage d'un message crypté à l'aide d'un algorithme de décalage.

1 Base de l'algorithme de cryptage

En cryptographie, le chiffrement par décalage, aussi connu comme le chiffre de César ou le code de César, est une méthode de chiffrement très simple utilisée par Jules César dans ses correspondances secrètes (ce qui explique le nom « chiffre de César »).

Le texte chiffré s'obtient en remplaçant chaque lettre du texte clair original par une lettre à distance fixe, toujours du même côté, dans l'ordre de l'alphabet. Pour les dernières lettres (dans le cas d'un décalage à droite), on reprend au début. Par exemple avec un décalage de 3 vers la droite, A est remplacé par D, B devient E, et ainsi jusqu'à W qui devient Z, puis X devient A etc. Il s'agit d'une permutation circulaire de l'alphabet.

La longueur du décalage, 3 dans l'exemple évoqué, constitue la clé du chiffrement qu'il suffit de transmettre au destinataire — s'il sait déjà qu'il s'agit d'un chiffrement de César — pour que celui-ci puisse déchiffrer le message. Dans le cas de l'alphabet latin, le chiffre de César n'a que 26 clés possibles (y compris la clé nulle, qui ne modifie pas le texte).

Le chiffre de César, seul, n'offre aucune sécurité de communication, à cause du très faible nombre de clés, ce qui permet d'essayer systématiquement celles-ci quand la méthode de chiffrement est connue, mais aussi parce que, comme tout encodage par substitution monoalphabétique, il peut être très rapidement « cassé » par analyse de fréquences (certaines lettres apparaissent beaucoup plus souvent que les autres dans une langue naturelle).

`https://fr.wikipedia.org/wiki/Chiffrement_par_décalage`

2 Programmation

2.1 Harmonisation du texte

Pour une harmonisation de la chaine d'entrée, écrire une fonction *supprime_interdit*, qui supprimera les caractères autres qu'alphabétiques, et transformera les majuscules en minuscules.

Exemple

chaine avant harmonisation : "J'aime manger des pommes."
chaine après harmonisation : "jaimemangerdespommes"

2.2 Cryptage

Ecrire une fonction *cryptage_chaine*, qui procède au cryptage d'une chaine, pour un décalage donné, en respectant les principes expliqués.

Exemple

chaine à cryter : "jaimemangerdespommes"
après un décalage de 2 :
chaine cryptée : "lckogocpigtfgurqoogu"

3 Decryptage

L'objectif de ce décryptage est d'essayer de comprendre le message, sans avoir la clé permettant de le faire avec certitude. Pour décrypter un tel message, il existe plusieurs solutions. Nous vous demandons de mettre en place les 2 propositions ci-dessous.

Dans les 2 cas, il sera nécessaire de mettre en place une fonction *decryptage_chaine* qui procède au décryptage d'une chaine, pour un décalage donné, en respectant les principes expliqués.

Exemple

chaine à décryter : "lckogocpigtfgurqoogu"
après un décalage de 2 :
chaine décryptée : "jaimemangerdespommes"

3.1 Méthode Brute

Cette première méthode se propose de tester tous les décalages possibles, jusqu'à ce que l'utilisateur soit satisfait du résultat.

Exemple

chaine à crypter : jaimemangerdespommes
chaine cryptée : lckogocpigtfgurqoogu
decalage de 0
chaine dé-crypté : lckogocpigtfgurqoogu
Le texte est-il comprenhensible ? 'o' ou 'n'
n
decalage de 1
chaine dé-crypté : kbjnfnbohfseftqpnnft
Le texte est-il comprenhensible ? 'o' ou 'n'
n
decalage de 2
chaine dé-crypté : jaimemangerdespommes
Le texte est-il comprenhensible ? 'o' ou 'n'
o

3.2 Méthode basée sur l'analyse de fréquence

L'analyse de fréquence consiste à examiner la fréquence des lettres employées dans un message chiffré. Cette méthode est fréquemment utilisée pour décoder des messages chiffrés par substitution, dont un exemple très simple est le chiffre de César.

Elle est basée sur le fait que, dans chaque langue, certaines lettres ou combinaisons de lettres apparaissent avec une certaine fréquence. Par exemple, en français, le e est la lettre la plus utilisée, suivie du a et du s. Inversement, le w est peu utilisé.

Ces informations permettent aux cryptanalystes de faire des hypothèses sur le texte clair, à condition que l'algorithme de chiffrement conserve la répartition des fréquences, ce qui est le cas pour des substitutions mono-alphabétiques et poly-alphabétiques.

`https://fr.wikipedia.org/wiki/Analyse_fréquentielle`

3.2.1 Programmation

— Ecrire une fonction *nb_occurrence*, qui calcule le nombre de fois où apparaît un caractère précis, dans une chaine.
Exemple
chaine : "jaimemangerdespommes"
nombre de 'a' : 2

— Ecrire une fonction *lettre_maximum_nb_occurrence*, qui renvoie la lettre qui apparaît le plus de fois dans le texte. *Exemple*
chaine : "lckogocpigtfgurqoogu"
lettre apparaissant le plus souvent : 'o'

— Modifier cette fonction *lettre_maximum_nb_occurrence*, qui renvoie la lettre qui apparaît le plus de fois dans le texte, en tenant compte de lettres déjà testées...(on enregistrera les lettres déjà testées dans une chaine)
Exemple
chaine : "lckogocpigtfgurqoogu"
lettre apparaissant le plus souvent : 'o'

chaine : "lckogocpigtfgurqoogu"
lettre apparaissant le plus souvent après le "o" : 'g'

chaine : "lckogocpigtfgurqoogu"
lettre apparaissant le plus souvent après le "og" : 'c'

chaine : "lckogocpigtfgurqoogu"
lettre apparaissant le plus souvent après le "ogc" : 'u'

Pour cela, on testera si la lettre courante est déjà présente dans la chaine des lettres testées. Si oui, on ne l'utilisera pas, sinon, elle sera une lettre potentielle au résultat de la fonction.

— Ecrire le programme principale, qui :

1. à partir d'une chaine cryptée, affiche la lettre apparaissant le plus souvent, en tenant compte des caractères déjà testée (cette chaine

est vide au départ !).

2. calcule la différence avec la lettre 'e', pour avoir une idée de décalage.
3. ajoute cette lettre à la liste des caractères déjà traitées
4. décrypte le texte et l'affiche
5. demande à l'utilisateur si le texte est compréhensible
6. si non, on recommence à l'étape 1
7. si oui, on a fini !

Exemple
chaine cryptée : lckogocpigtfgurqoogu
Lettre apparaissant le plus souvent : o
chaine décryptée : bsaewesfywjvwkhgeewk
Le texte est-il compréhensible ? 'o' ou 'n'
n
chaine cryptée : lckogocpigtfgurqoogu
Lettre apparaissant le plus souvent après o : g
chaine décryptée : jaimemangerdespommes
Le texte est-il compréhensible ? 'o' ou 'n'
o
Bravo !