UPMC SORBONNE UNIVERSITÉS

Cryptanalyse du chiffrement de Vigenère en pratique

Version du 26 janvier 2015

TME

Choix du langage

Le langage choisi doit permettre une maîtrise des éléments suivants
- Entrées/sorties standard, fichier, arguments de la ligne de commande.
- Parcours de chaîne de caractères et calcul arithmétique sur les caractères, transformation minuscule/majuscule.

À titre d'exemple nous ferons des rappels pour le langage C ou le langage Python.

Exercice 1 – Échauffement

1. Écrire un programme permettant de calculer la table de multiplication modulo un entier n donné en entrée.
2. En utilisant la fonction de la question précédente, écrire un programme permettant de calculer l'inverse de m modulo un n deux entiers donnés en entrée.
3. Écrire un programme permettant de calculer l'ensemble des diviseurs (positifs) d'un entier n donné en entrée. Pour ceux qui utilisent Python, faites cela en construisant une unique liste.
4. En utilisant la fonction de la question précédente, écrire un programme permettant de retrouver le plus grand diviseur commun de deux entiers m et n donnés en entrée.

Exercice 2 – Fréquence des lettres dans une langue donnée

1. Écrire un programme `frequence` qui étant donné un nom de fichier texte codé en ASCII comme argument analyse la fréquence d'apparition de chacun de ses caractères.
   Il s'agit essentiellement d'écrire une fonction qui remplit un tableau de correspondance caractère $\longrightarrow$ fréquence sous forme de pourcentage (on pourra commencer par renvoyer le nombre d'apparition de chaque lettre et le nombre total de lettres). Cette fonction sera réutilisée dans les exercices et séances suivantes pour la cryptanalyse de textes. Ici on illustre cette fonction avec un `main` qui affiche le tableau.
2. Établir un histogramme des fréquences pour des textes écrits en français, en anglais ou tout autre langue (on pourra consulter le site web).
   Ici il s'agit d'établir des tableaux de fréquences de référence pour différentes langues, qui nous serviront de comparaison par la suite pour la cryptanalyse de textes.
   Pour ce faire, vous utiliserez la fonction de la question précédente sur un texte d'une langue donnée. Ces textes seront supposés écrits sans caractère accentué (i.e. en ASCII de base) et suffisamment longs. Pour produire de tels fichiers texte, vous pourrez utiliser les sites web `http://abu.cnam.fr/index.html` pour des textes en français exclusivement ou `http://www.gutenberg.org/wiki/Main_Page` Vous pourrez utiliser le script `nettoie` (fourni dans le répertoire de l'UE) pour éliminer les accents et normaliser le texte.
   Afin de représenter vos sorties de manière graphique (comme en cours), vous utiliserez le logiciel GNUplot (voir `http://www.gnuplot.info/documentation.html` ou `?histograms` sous l'interprète `gnuplot`) sur la sortie de votre programme. Le script `dessine_histogramme` (fourni dans le répertoire de l'UE) vous permettra d'afficher un tel histogramme lorsque le résultat de votre commande `frequence` est bien formaté (lettre sur la première colonne suivie d'un espace et du nombre de fois qu'elle apparaît dans le fichier texte).

Exercice 3 – Chiffrements et déchiffrements

1. Écrire un programme permettant de réaliser le chiffrement et déchiffrement de César.
   Vous spécifierez bien votre programme et vérifierez son comportement sur les exemples de l'exercice 4 de la première feuille.
2. Écrire un programme permettant de réaliser un chiffrement et déchiffrement mono-alphabétique.
   Vous spécifierez bien votre programme et vérifierez son comportement sur les exemples de l'exercice 4 de la première feuille.
3. Écrire un programme permettant de réaliser le chiffrement et déchiffrement de Vigenère.
   Vous spécifierez bien votre programme et vérifierez son comportement sur les exemples de l'exercice 5 de la première feuille.

Exercice 4 – Cryptanalyse d'un chiffrement mono-alphabétique

1. Écrire un programme permettant de réaliser la cryptanalyse d'un texte en français chiffré à l'aide d'un chiffrement mono-alphabétique.
   Utiliser la fonction de calcul des fréquences des caractères dans un texte pour cryptanalyser des chiffrements mono-alphabétiques.
   Vous essaierez dans un premier temps d'utiliser cette fonction comme une aide à une cryptanalyse manuelle. Vous pourrez ensuite développer un programme interactif qui propose les meilleurs substitutions possibles et laisse à l'utilisateur le choix des remplacements successifs.
   Pour vous entraîner, utiliser les fichiers chiffrés disponibles sur le site web de l'UE.
2. Automatiser au mieux cette cryptanalyse en utilisant la corrélation de Pearson.
3. En utilisant le programme `monobi` du répertoire de l'UE `/Infos/lmd/2013/licence/ue/li336-2014fev/`, ou celui que vous aurez programmé aux questions précédentes, cryptanalyser les textes fournis dans ce répertoire.
   Le programme `monobi` permet interactivement de chiffrer / déchiffrer ou cryptanalyser un chiffrement mono-alphabétique en fournissant les fréquences du texte, celles de la langue pour comparaison et la possibilité de substituer les correspondances petit à petit : le texte chiffré est écrit en majuscules, on transforme le chiffré majuscule en déchiffré minuscule. Les explications sont fournis au fur et à mesure du déroulement. Pour faciliter la cryptanalyse cette version utilise la fréquence des bigrammes, c'est-à-dire de couples de lettres dans le texte à cryptanalyser et dans le texte de référence de la langue.

Exercice 5 – Indice de Coïncidence vs Vigenère

1. Écrire une fonction permettant de calculer l'indice de coincidence d'un texte.
2. Écrire une fonction permettant de calculer la longueur de la clé d'un chiffrement de Vigenère.
3. En utilisant la cryptanalyse de chiffrement par décalage, finir de cryptanalyser un chiffrement de Vigenère une fois que la longueur de la clé est trouvée.
4. Pour les plus avancés, vous pourrez implanter un programme de cryptanalyse du chiffrement de Vigenère comme vous l'avez fait pour le chiffrement par substitution. Vous proposerez l'utilisation de la corrélation de Pearson ou l'indice de multi-coïncidence à l'utilisateur pour pouvoir en faire la comparaison.

