Cryptographie

TD n°1 : Introduction & Histoire de la cryptographie

Partie I

1 Chiffrement Caesar

En utilisant la substitution A ⟶ E, B ⟶ F, ..., chiffrez le texte suivant:

```
ATTAQUEZ AU CREPUSCULE
```

2 Avé Caesar !

Le texte suivant a été chiffré suivant un chiffrement de substitution de type Caesar. Saurez-vous retrouver le message originel ?

```
WR AR CRAFR CNF DH VY L NVG QR OBAAR BH QR ZNHINVFR FVGHNGVBA
```

Remarque: vous pouvez écrire un programme en Python pour vous aider

Partie II

On a intercepté le texte suivant:

```
VLETDIBTLRVWISIDIZSFDJAAQHEKXVTWMXWQRMILMFJVEBWRVZJUZDLWWSELWZSSKMIEEAXFNAGYKQVNWJQWU
CABSRWVLITSGGZMRMDETWZTMWXDTDBZSKWYFUBWRLNVRMBEARVCMHEQXVTRSMEJPVQNVWURRKWUMJAYIAQWXI
AKARWXTEAHVGKFNLWWWYUETAGSYVSZSGASVSYMMSYKAKZIFYLNPGQEJRLILIJWVOCKSFYEEAWXETITAKIKFZE
CPUMNCABLEUMVNBSGWVLOVHIFXVEBSGWVLOVEEFLVACPYKFXEAUSERVACPRGZIRQLYJJJACPPGHLTQGRKQFCI
DIKFLXQFXGSRTQGRKIVSXSCKFESIMBGIVUZKHMXFLLWWNNCLIYIKJKDMDEAWCUQEIEJAAQEIEFDAQKSFTLJIA
KJFEDQVIEJJFMFILWVSRWZGNJLIKIASVQCAGGZCETWPGSXDMESFORRLARVJIRQWVWQRRWMXWUIEAIYWHYEHES
AQRGZSRVJVTTSVYJJEQFIIZZVIVIJTLEVSYZFMRMUSMAVRBWHWGRTMSYPVLIXSWKJET
```

Grâce à votre réseau d'espions, vous savez qu'il s'agit d'un texte clair en français qui a été chiffré avec un système de Vigenère, réputé inviolable ! Cela ne va cependant pas vous arrêter, et vous vous attelez à la lourde tâche de déchiffrer ce message cryptique...

1 Décryptage automatique du chiffre Caesar

Dans la partie précédente, vous avez probablement essayé les 25 chiffrements Caesar différents pour voir "à l'oeil" quel déchiffrement produit un texte français lisible. C'est une solution, mais on peut faire mieux : retrouver directement le chiffrement utilisé grâce à la méthode des fréquences.

En français, les fréquences des lettres sont les suivantes :

```
frequences = {'A': 0.0815, 'N': 0.0712, 'B': 0.0097, 'O': 0.0528, 'C': 0.0315, 'P':
0.028, 'D': 0.0373, 'Q': 0.0121, 'E': 0.1739, 'R': 0.0664, 'F': 0.0112, 'S': 0.0814,
'G': 0.0097, 'T': 0.0722, 'H': 0.0085, 'U': 0.0638, 'I': 0.0731, 'V': 0.0164, 'J':
0.0045, 'W': 0.0003, 'K': 0.0002, 'X': 0.0041, 'L': 0.057, 'Y': 0.0028, 'M': 0.0287,
'Z': 0.0015}
```

Afin de retrouver automatiquement le bon texte clair pour un texte chiffré par un chiffrement Caesar, vous pouvez appliquer l'algorithme suivant :

- Appliquez un déchiffrement Caesar (B ⟶ A puis C ⟶ A etc)
- Calculez la fréquence des lettres du nouveau texte obtenu
- Effectuez un produit scalaire entre le vecteur des fréquences du texte et le vecteur de fréquence des lettres dans la langue française

- Le texte qui maximise ce produit scalaire a de fortes chances d'être le bon texte clair

Implémentez une telle fonction de déchiffrement automatique. Fonctionne-t-elle sur le texte chiffré de la partie I ? Pourquoi cela pourrait-il ne pas être le cas ?

Question bonus Pourquoi cela fonctionne-t-il ?

2 Détermination de la taille de la clé utilisée

Armé de cet algorithme, nous allons pouvoir déterminer la taille de la clé utilisée pour le chiffrement Vigenère. En effet, on peut remarquer que pour une clé de taille L, le texte obtenu en ne gardant qu'une lettre sur L du chiffré est en fait le chiffrement, par un chiffre de type Caesar, du texte clair dans lequel on n'aurait gardé qu'une lettre sur L.

Ainsi, si on a correctement déterminé la longueur de la clé, alors l'algorithme de la question précédente devrait trouver un déchiffrement Caesar associé à un produit scalair important (pour le chiffré obtenu en gardant une lettre sur L), alors que pour une longueur de clé incorrecte, tous les produits scalaires devraient être relativement faibles.

Sachant que la clé est de longueur inférieure à 20, déterminez celle-ci. Vous pouvez sélectionner le sous-texte obtenu un gardant une lettre sur L en Python en utilisant la syntaxe `texte[::L]`. Pour toutes les hypothèses de taille L entre 1 et 20, calculer le produit scalaire le plus important en appliquant tous les déchiffrement Caesar possibles, et affichez ces valeurs. Que remarquez-vous ?

3 Décryptage complet

Une fois la taille de la clé obtenue, il devrait être aisé de décrypter entièrement le texte en effectuant L décryptage Caesar successifs. Quelle était la clé ?

Partie III - Challenge

Nous sommes en 1945. Les forces alliées ont préparé depuis des mois un débarquement massif en Normandie. La date approche, mais vous voulez être sûrs que l'Axe n'est pas au courant de vos plans. En effet, de multiples campagnes d'intox ont eu lieu pour brouiller les pistes et tromper les services de renseignement ennemis.

Enfin, un beau jour, vous interceptez un message de l'occupant allemand en France. Vous savez que ce message contient le lieu dans lequel les Allemands pensent que le débarquement aura lieu. Voici le message intercepté :

```
LBKGBMNVRPMVBNOAXKJWCQWARXPFGTXAVBIWVCGEEXWMGLVFDGYDXMGFGVTTCGXIRSTPEEUTRDNPMTBNPPE
WYWINJEAFEHPZLCMWWRUDCVRTAHQAHDATXWWKTPRTRHXKLBDIAEGPZYGZXIROCIMHWHVDRXIVLSSGMTXLSZQTWFQME
```

Les informations que vous possédez concernant ce message sont les suivantes :
- Il s'agit d'un texte clair en **français**
- Il a été chiffré avec un chiffre de Vigénère, avec une clé aussi longue que le message initial
- La clé utilisée est issue d'un texte en **français**
- Le lieu du débarquement est un nom de ville, plus précisément, une ville parmi les trois suivantes :
  - Cherrueix
  - Pignochet
  - Deauville

L'objectif est de déterminer le lieu prévu du débarquement, et justifier la réponse. En particulier, vous devrez expliquer la démarche suivie pour parvenir à vos fins et indiquer à quelle position du message intercepté se trouve le nom de la ville prévue par les Allemands.

Vous pouvez tentez ce challenge par petits groupes (maximum 3 personnes). Si vous arriver à fournir la bonne réponse avant la fin de l'année 2025, cela pourra constituer le sujet de votre restitution finale avec une note minimale garantie de 15/20.

Vous pouvez vous aider pour ce challenge de deux fichiers préparés pour l'exercice, présents dans le répertoire du cours :

- Un premier qui contient la concaténation normalisée (tout en majuscule, sans espaces, accents etc.) de plusieurs longs textes français (mais l'extrait utilisé pour chiffrer ne se trouve **pas** dans ce corpus !), nommé *textes_fr.txt*
- Un second qui contient la liste des mots du dictionnaire français, normalisés de la même façon, nommé *mots_fr.txt*

Bonne chance !