Scripting Shell et Python

TP1 : Chiffrement et décryptage - Sujet

Préliminaires

Dans ce TP, on s'intéresse au chiffrement de textes par substitution, à leur déchiffrement, et enfin au moyen de *casser* un tel cryptage. Un chiffrement par substitution consiste simplement à remplacer systématiquement une lettre par une autre.

Certains chiffrements par substitution sont dits *par décalage* ou appelés *chiffre de César* : dans ces cas, une lettre et sa remplaçante sont toujours séparées dans l'alphabet par le même nombre de lettres. Un chiffrement par décalage répandu sur le web se nomme ROT13 : une lettre et sa remplaçante sont à une distance de 13 dans l'alphabet (le *a* est remplacé par le *n*, le *b* par le *o*, etc.).

Par souci de simplicité, on ne considère que des textes en minuscules, sans accent et sans symbole de ponctuation. Seul l'espace est utilisé entre les mots mais n'est pas remplacé par le chiffrement.

Ce TP va nécessiter de traiter des chaînes de caractères qu'il peut être commode de voir comme des listes de caractères, ce que permet Python.

Concernant les caractères eux-mêmes, on rappelle qu'il est possible de repérer un caractère par un numéro (son code ASCII).

Nous aurons également besoin de stocker les correspondances entre lettres, ce qui peut être fait à l'aide de *dictionnaires Python*.

Enfin, pour décrypter un texte sans connaître la règle de chiffrement, il est courant d'utiliser la fréquence d'apparition des lettres dans la langue choisie. On considérera qu'en français les lettres se rangent comme suit, de la plus fréquente à la moins fréquente :

e a i t s n l u r o d m c p v q h f b g j x y w z k

Fonctions à implémenter

Il faut d'abord réaliser et tester quelques fonctions pour gérer les dictionnaires, chiffrer et déchiffrer des textes.

1. *chiffrement_lettre(l, d)* renvoie la correspondance de la lettre *l* dans le dictionnaire *d* si cette correspondance existe, renvoie la lettre *l* elle-même sinon.
2. *chiffrement_phrase(p, d)* construit une nouvelle chaîne de caractères correspondant au chiffrement caractère par caractère de la phrase *p* à l'aide du dictionnaire *d*. Définir un dictionnaire et tester ces fonctions en codant une phrase quelconque.
3. *inverse_dico(d)* renvoie un nouveau dictionnaire qui inverse les clefs et les valeurs du dictionnaire *d*. Tester en déchiffrant la phrase précédemment codée.
4. *dico_rot_13 ()* construit un dictionnaire correspondant au chiffrement en *ROT13*. Tester à nouveau pour chiffrer une phrase quelconque. Quelles solutions sont possibles pour le déchiffrement ?

On veut maintenant déchiffrer un texte codé par une technique de substitution mais on ne dispose pas du dictionnaire utilisé. L'objectif est de décoder le texte mystère suivant (voir texte à décoder).

Les étapes à réaliser sont les suivantes

5. *compte_lettres(p)* construit un dictionnaire faisant correspondre chaque lettre apparaissant dans la phrase *p* à son nombre d'occurrences dans cette même phrase. Tester sur le texte mystère.
6. *tri_bulles_dico (d)* est un tri à bulles modifié pour renvoyer les clefs d'un dictionnaire *d*, ordonnées par valeurs décroissantes. Tester sur le dictionnaire précédemment calculé par *compte_lettres*.
7. *arrays2dict(ks, vs)* renvoie un dictionnaire dont les clefs correspondent au tableau *ks* et qui associe pour chacune de ces clefs la valeur se trouvant à la même position dans le tableau *vs*. Utiliser cette fonction pour combiner le tableau fourni par *tri_bulles_dico* et le tableau des lettres de l'alphabet classées par fréquences décroissantes dans la langue française.
8. *decrypte (pc, ll)* doit décrypter la phrase pc à l'aide des lettres de l'alphabet rangées par ordre de fréquence décroissante dans la langue utilisée et disponible dans le tableau *ll*. Décoder le texte mystère à l'aide de cette fonction.

Texte à décoder

or z f kcgrkcgh fnnggh mg ug rofo onaougugna fqgb cn u eorrofu rgswfny or gafoa y cng fnbognng xfuorrg iwpaghafnag ga mfyoh or fqfoa gag woblg ufoh cng hgwog yg ufrlgcwh r fqfoa wgycoa f rf uohgwg ipcw gqoagw r lcuorofaopn yg hgh yghfhawgh or kcoaaf rf npcqgrrg pwrgfnh rf qorrg yg hgh fogcj ga gafdroa hf ygugcwg yfnh r org yg hcrroqfn iwgh blfwrghapn yfnh rf bfwprong yc hcy bgaag org gha ygh irch honscrogwgh grrg n gha scgwg bpuiphgg kcg yg hfdrg yg ugw ga f gnqowpn awpoh uorrgh yg rpns gn rfwsgcw grrg n f mfufoh irch y cn kcfwa yg uorrg grrg gha hgifwgg yc bpnaongna ifw cng bwokcg f igong qohodrg kco xorawg f awfqgwh cng ufhhg yg wphgfcj ga yg qfhg wgnygt qpch lfdoacgr ygh ipcrgh y gfc rf qgsgafaopn bpuug pn igca rg hciiphgw gha ifcqwg pc ipcw fonho yowg nfong pn n z awpcqg ifh y fwdwgh y cng bgwafong yougnhopn qgwh r gjawguoag pbboygnafrg f r gnywpoa pc h grgqgna rg xpwa upcrawog ga kcgrkcgh uohgwfdrgh dfaohhgh yg dpoh lfdoaggh ignyfna r gag ifw rgh sgnh kco xcogna rgh ipchhogwgh ga rgh xogqwgh yg blfwrghapn pn wgnbpnawg or gha qwfo rg ifruogw nfon hgaosgwg ufoh apcag r org f r gjbgiaopn yg bg ipona pbboygnafr ga y cn ghifbg awohag ga drfnblfawg kco dpwyg rf ugw gha bpcqgwag y gifohhgh dwpchhforrgh yg uzwag pypwoxgwfna ho ghaoug ifw rgh lpwaobcragcwh fnsrfoh r fwdchag z upnag hpcqgna f cng lfcagcw yg kcontg pc qonsa iogyh or z xpwug cn aforroh iwghkcg ouigngawfdrg ga blfwsg r fauphilgwg yg hgh ifwxcuh fc irch iwpxpny yg bg aforroh npn rpon yg r gjawguoag pwognafrg yg r org b gha f yowg yg rf irch grposngg rgswfny h gafoa dfao rco ugug cng igaoag lcaag kc or pbbcifoa kcfny ipcw rf iwguogwg xpoh ga ifw lfhfwy mg xoh hf bpnnfohhfnbg bgaag bpnnfohhfnbg ucwoa dogn qoag gn fuoaog bfw or z fqfoa bgwagh yfnh rg blgw wgbrch yg kcpo gjboagw r onagwga ga r ghaoug mg qoh kc or fqfoa wgbc cng xpwag gycbfaopn lgcwgchgugna hgwqog ifw ygh xfbcragh hiowoacgrrgh igc bpuucngh ufoh kc or gafoa onxgbag yg uohfnalwpiog ga hcmga f yg ufrlgcwgchgh fragwnfaoqgh y gnalpchofhug ga yg ugrfnbprog dogn kc or gca blgt rco dgfcbpci yg roqwgh or h gn hgwqfoa wfwgugna hgh iwonboifcj fuchgugnah bpnhohafogna f blfhhgw ga f igblgw pc f xrfngw hcw rf irfsg ga f awfqgwh rgh uzwagh gn kcgag yg bpkcorrfsgh ga y gblfnaorrpnh gnapuprpsokcgh hf bprrgbaopn fcwfoa ic xfowg gnqog f cn hefuugwyfu yfnh bgh gjbcwhopnh or gafoa pwyonfowgugna fbbpuifsng ifw cn qogcj ngswg npuug mcioagw kco fqfoa gag fxxwfnblo fqfna rgh wgqgwh yg rf xfuorrg ufoh kc pn n fqfoa ic ygboygw no ifw ugnfbgh no ifw iwpughhgh f fdfnypnngw hpn mgcng ufhhf eorr or bpnhoygwfoa bpuug hpn ywpoa yg rg hcoqwg ifwapca or n gha ifh ouiwpdfdrg kcg rgh ifwgnah yg rgswfny mcsgfna kcg bgrco bo fqfoa rf agag cn igc ygwfnsgg hg hpogna fiirokcgh f bpnxowugw mcioagw yfnh hpn pdhaonfaopn yfnh rg dca yg ugaawg cng ghigbg yg sfwyogn ga yg hcwqgorrfna fciwgh yc xcsoaox hpch rf rfaoacyg yg r org yg hcrroqfn rgh loqgwh hpna wfwgugna wospcwgcj ga b gha cn gqgngugna kcfny fc ygbron yg r fnngg rg xgc ygqogna onyohignhfdrg bgignyfna qgwh rg uorogc y pbapdwg or z gca cng mpcwngg y cn xwpoy wgufwkcfdrg mchag fqfna rg bpcblgw yc hprgor mg ug xwfzfoh cn blguon f awfqgwh rgh aforroh qgwh rf lcaag yg upn fuo kcg mg n fqfoh ifh qc ygicoh kcgrkcgh hgufongh mg ygugcwfoh frpwh f blfwrghapn f cng yohafnbg yg ngcx uorrgh yg r org ga rgh xfboroagh ipcw frrgw ga wgqgnow gafogna dogn uponh swfnygh kc fcmpcwy lco gn fwwoqfna f rf lcaag mg xwfiifo hgrpn upn lfdoacyg ga ng wgbgqfna ifh yg wgipnhg mg blgwblfo rf brgx pc mg hfqfoh kc grrg gafoa bfblgg m pcqwoh rf ipwag ga m gnawfo cn dgfc xgc xrfudfoa yfnh rg xpzgw b gafoa cng hcwiwohg ga f bpci hcw cng ygh irch fswgfdrgh mg ug ygdfwwfhhfo yg upn ifrgapa mg awfonfo cn xfcagcor fciwgh ygh dcblgh igaorrfnagh ga m faagnyoh ifaoguugna r fwwoqgg yg ugh lpagh igc fiwgh rf apudgg yg rf ncoa orh fwwoqgwgna ga ug xowgna cn fbbcgor apca f xfoa bpwyofr mcioagw apca gn wofna y cng pwgorrg f r fcawg hg ypnnfoa yc upcqgugna ga iwgifwfoa kcgrkcgh ipcrgh y gfc ipcw rg hpcigw rgswfny gafoa yfnh cng yg hgh bwohgh y gnalpchofhug bfw yg kcgr fcawg npu fiigrgw bgrf or fqfoa awpcqg cn doqfrqg onbpnnc xpwufna cn sgnwg npcqgfc ga uogcj gnbpwg or fqfoa blfhhg ga faawfig fqgb r fhhohafnbg yg mcioagw cn hbfwfdgg kc or bwpzfoa apca f xfoa npcqgfc ga hcw rgkcgr or yghowfoa fqpow upn pionopn rg rgnygufon ufaon