Substitution Ciphers

Substitution cipher is one of the oldest and simplest. To use it, Alice and Bob share a *private key* that represents a permutation of the letters, e.g.,

jsuyfhkpicomxrqatlbvznewgd

To encode her message, Alice starts by aligning the key with the alphabet:

abcdefghijklmnopqrstuvwxyz

jsuyfhkpicomxrqatlbvznewgd

Then, she uses the resulting substitution table as follows: she replaces every letter `a` in her message by `j`, every `b` by `s`, and so on. For decoding, one uses the same substitution table with the two rows switched: `j` is replaced by `a`, `s` is replaced by `b`, and so on. It is particularly easy to implement this cipher in python:

```
alphabet = 'abcdefghijklmnopqrstuvwxyz'
key =      'jsuyfhkpicomxrqatlbvznewgd'


def substitute(text, substitute_what, substitute_by):
    result = ''
    for symbol in text.lower():
        if symbol in substitute_what:
            result += substitute_by[substitute_what.index(symbol)]
        else:
            result += symbol

    return result


def encode(plaintext):
    return substitute(plaintext, alphabet, key)


def decode(ciphertext):
    return substitute(ciphertext, key, alphabet)


message = 'the quick brown fox jumps over the lazy dog'
code = encode(message)
print(code)
print(decode(code))
```

Run

Reset

```
vpf tziuo slqer hqw czxab qnfl vpf mjdg yqk
```

"The quick brown fox jumps over the lazy dog" is a well-known *pangram*, that is, a sentence containing all letters of the alphabet. For this reason, it is widely used for testing fonts.

A special case of the substitution cipher where the key is a *cyclic shift* of the alphabet is known as *Caesar cipher*. It is named after Julius Caesar, who used it for establishing secure communication. A key for such an encryption scheme can be generated using a physical device like the one shown below:

Two rotating disks for generating a cyclic shift of the alphabet. Source: Wikipedia.

This is how one can generate a cyclic shift in python:

```
alphabet = 'abcdefghijklmnopqrstuvwxyz'
key = alphabet[3:] + alphabet[:3]

print(alphabet)
print(key)
```

Run

Reset

```
abcdefghijklmnopqrstuvwxyz
defghijklmnopqrstuvwxyzabc
```

In this example, every letter is replaced by a letter three places further in the alphabet. Nowadays, it is not recommended to use Caesar cipher as it is too easy to crack it.

Problem

The following ciphertext is obtained using Caesar cipher:

vpf hinf sqwirk eidjlyb czxa tziuomg

Try to decode it.

The reason why this cipher is easily breakable is that *the space of possible keys is small*. Indeed, there are only 26 different cyclic shifts of the alphabet: the possible values of the shift are $0, 1, \ldots, 25$. This makes it possible for Eve to enumerate all the keys and to decode using each of them.

```
ciphertext = 'kyv wzmv sfozex nzqriuj aldg hlztbcp'
for shift in range(26):
    key = alphabet[shift:] + alphabet[:shift]
    print(decode(ciphertext))
```

```
kyv wzmv sfozex nzqriuj aldg hlztbcp
jxu vylu renydw mypqhti zkcf gkysabo
iwt uxkt qdmxcv lxopgsh yjbe fjxrzan
hvs twjs pclwbu kwnofrg xiad eiwqyzm
gur svir obkvat jvmneqf whzc dhvpxyl
ftq ruhq najuzs iulmdpe vgyb cguowxk
esp qtgp mzityr htklcod ufxa bftnvwj
dro psfo lyhsxq gsjkbnc tewz aesmuvi
cqn oren kxgrwp frijamb sdvy zdrltuh
bpm nqdm jwfqvo eqhizla rcux ycqkstg
aol mpcl ivepun dpghykz qbtw xbpjrsf
znk lobk hudotm cofgxjy pasv waoiqre
ymj knaj gtcnsl bnefwix ozru vznhpqd
xli jmzi fsbmrk amdevhw nyqt uymgopc
wkh ilyh eralqj zlcdugv mxps txlfnob
vjg hkxg dqzkpi ykbctfu lwor swkemna
uif gjwf cpyjoh xjabset kvnq rvjdlmz
the five boxing wizards jump quickly
sgd ehud anwhmf vhyzqcr itlo pthbjkx
rfc dgtc zmvgle ugxypbq hskn osgaijw
qeb cfsb ylufkd tfwxoap grjm nrfzhiv
pda bera xktejc sevwnzo fqil mqeyghu
ocz adqz wjsdib rduvmyn ephk lpdxfgt
nby zcpy vircha qctulxm dogj kocwefs
max ybox uhqbgz pbstkwl cnfi jnbvder
lzw xanw tgpafy oarsjvk bmeh imaucdq
```

In the output, one identifies the original message easily: only one of them (starting with `the`) consists of English words.

When the key is not just a cyclic shift of the alphabet, but rather an arbitrary permutation of the alphabet, enumerating all keys is not that easy: the size of the space of all keys is

$$26! = 403\,291\,461\,126\,605\,635\,584\,000\,000.$$

Still, substitution ciphers are not considered secure as they are vulnerable to frequency analysis attacks. Such attacks exploit the fact that some letter combinations appear more frequently in English texts than others: the letters `e`, `t`, `a`, `o` are the most common; the word `the` usually appears many times.

Problem

Try decoding the following ciphertext, taken from "The Gold-Bug" by Edgar Allan Poe. (All white spaces and punctuation is removed.)

livitcswpiyvewhevsriqmxleyveoiewhrxeipfemvewhkvstylxzixlikiixpijvszeyperrgerimwqlmglmxqeriwgpsrihmxqerekietxmjtprgevekeitrewhexxlexxmzitwawsqwxswextvepmrxrsjgstvrieyviexcvmuimwergmiwxmjmgcsmwxsjomiqxliviqivixqsvstwhkpegarcsxrwievswiibxvizmxfsjxlikegaewhepswyswiwievxlisxlivxlirgepirqiviibgiihmwypflevhewhypsrrfqmxleppxlieccievewgisjktvwmrlihysphxliqimylxsjxlimwrigxqeroivfvizevaekpiewhxeamwyeppxlmwyrmwxsgswrmhivexmswmgstphlevhpfkpezintcmxivjsvlmrscmwmswvircigxmwymx

See Wikipedia for a hint.

Completed **Go to next item**

Like **Dislike** **Report an issue**