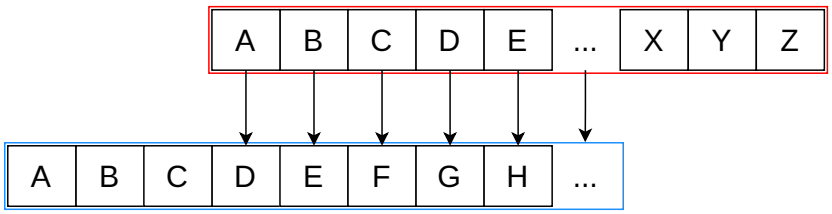
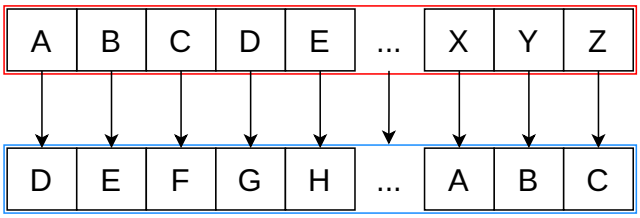
0.1 Cifrari Antichi - Cesare

Cominciamo ad analizzare i primi cifrari e iniziamo dai primi cifrari mai inventati, infatti come avevamo gia detto nell'introduzione il **Cifrario di Cesare** risale all'impero romano. Il cifrario fu inventato per inviare ai legionari informazioni per proseguire la battaglia. Chiaramente non potevano mandarli i messaggi in chiaro, anche perchè se gli avversari ne venivano a conoscenza potevano rispondere con una contromossa. Per questo **Giulio Cesare** invento un cifrario che si basava sullo spostamento di un determinato numero di posizioni le lettere nell'alfabeto. Per esempio il classico esempio di cifrario di cesare è di **Spostare di 3 posizioni in dietro**.

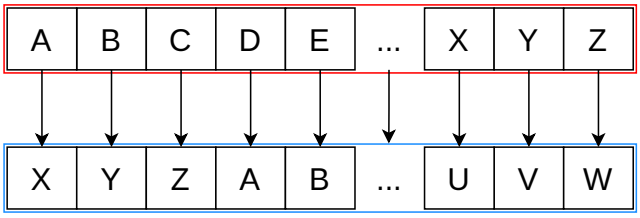


Le lettere che "escono a sinistra" le associamo alle ultime 3 lettere



Per cifrare un messaggio basta prendere ogni lettera del messaggio e trasformarlo nella lettere associata nella tabella. Ad esempio se volessi convertire la parola **Abbecce**, con il cifrario di cesare sarebbe **Deehffh**.

Per decifrarlo invece basta fare lo stesso procedimento ma l'alfabeto va spostavo verso destra, quindi con questo esempio il cifrario di decifratura sarà



In questo esempio abbiamo spostato l'alfabeto di 3 posizioni, ma questo è un numbero che possiamo scegliere noi a nostro piacimento. Secondo la tradizione Giulio Cesare usava il numero 3, però si possono anche spostare di quanto si vuole. Visto che questo numero è decisivo per la cifratura, il numero di spostamnenti non è altro che la **chiave** del sistema. Chiaramente ad una chiave diversa c'è un cifrario diverso, e di conseguenza una cifratura dei messaggi diversa.

Una possibile implementazione in python della cifrario di cesare a 3 spostamenti:

```
def Cifratura_Cesare(mess):
    return "".join([chr(((ord(char) - ord('a') + 3) % 26) +
        ord('a')) for a in mess])

def Decifratura_Cesare(mess):
    return "".join([chr(((ord(char) - ord('a') - 3) % 26) +
        ord('a')) for a in mess])
```

Online questa cifratura viene anche chiamata **ROT** (abbreviazione di *ROTation*) affiancato dal numero si spostamenti. Quindi **ROT3** sarà con 3 spostamenti (come gli esempi di prima) mentre **ROT15** sposterà di 15 caselle. In python un'implementazione per un generico **ROT-n**

```
def Cifratura_ROTn(mess, n):
    return "".join([chr(((ord(char) - ord('a') + n) % 26) +
        ord('a')) for a in mess])

def Decifratura_ROTn(mess, n):
    return "".join([chr(((ord(char) - ord('a') - n) % 26) +
        ord('a')) for a in mess])
```

Quindi se qualcuno vuole usare questo cifrario basta che decida un numero tra 1 e 25 e tenerserlo per se stesso. Se non fosse che questo metodo è soggetto ad attacchi **brute-force** infatti come abbiamo detto questo cifrario ha al massimo **25 chiavi possibili**, quindi qualcuno potrebbe creare un programma che provi tutte le possibili combinazioni. Un possibile programma in python:

```
def attacco_rotn(messDaDecifrare):
    for i in range(1, 26):
        print(Decifratura_ROTn(messDaDecifrare, i))
```

0.1.1 Cifrari Monoalfabetici

In realtà l'attacco brute force non è l'unico al quale è soggetto questo cifrario. Questo tipo di cifrario è definito **Monoalfabetico** perchè la stessa lettera viene sempre cifrato con la stessa lettera corrispondente. Tutti i cifrari Monoalfabetici sono soggetti ad attacchi detti **Letter Frequency**. Questo vuol dire che la **percentuali** di una lettera nel testo originale è la stessa della lettera cifrata nel testo cifrato. Facciamo un esempio con **Abbcccdddd**, che cifrato con ROT3 diventa **Deefffgggg**

Abbcccdddd $\Longrightarrow$ Deeffgggg

Lettera	**percentuale**
A	10%(1/10)
B	20%(2/10)
C	30%(3/10)
D	40%(4/10)

Lettera	**percentuale**
D	10%(1/10)
E	20%(2/10)
F	30%(3/10)
G	40%(4/10)

Questa caratteristica dei sistemi Monoalfabetici può essere sfruttata per decifrare, anche parzialmente, il messaggio. Questo perche noi possiamo creare delle tabelle con la frequenza di ogni lettera per ogni lingua. Per esempio la **tabella di frequenza** della lingua italiana è

Lettera	Frequenza
E	11.49 %
A	10.85 %
I	10.18 %
O	9.97 %
N	7.02 %

...

link alla tabella completa :
https://www.sttmedia.com/characterfrequency-italian

Questa tabella vuol dire che di media, un messaggio/parola in italiano ha quelle probabilità che contenga quelle lettere. Quindi perchè stiamo facendo tutto questo discorso? perchè se noi facciamo la stessa analisi anche sul messaggio cifrato e paragoniamo le lettere più presenti nel messaggio cifrato con la tabella sopra può essere che qualche lettera la troviamo. Chiaramente più è lungo il messaggio più è probabile di indovinare le lettere.

N.B. Chiaramente esistono tabelle di frequenza per ogni lingua, io qua ho portato quella italiana come esempio.

Facciamo un esempio pratico: supponiamo di avere questo messaggio:

hlvc irdf uvc crxf uz Tfdf, tyv mfcxv r dvqqfxzfief, kir ulv trkvev efe zekviifkkv uz dfekz, klkkf r jvez v r xfcwz, r jvtfeur uvccf jgfixviv v uvc izvekiriv uz hlvccz, mzve, hlrjz r le kirkkf, r izjkizexvijz, v r giveuvi tfijf v wzxlir uz wzldv, kir le gifdfekfizf r uvjkir, v le'rdgzr tfjkzvir urcc'rckir grikv; v zc gfekv, tyv zmz tfexzlexv cv ulv izmv, gri tyv iveur retfi gzu jvejzszcv rcc'fttyzf hlvjkr kirjwfidrqzfev, v jvxez zc glekf ze tlz zc crxf tvjjr, v c'Ruur izetfdzetzr, gvi izgzxczri gfz efdv uz crxf ufmv cv izmv, rccfekrereufjz uz elfmf, crjtzre c'rthlr uzjkveuvijz v irccvekrijz ze elfmz xfcwz v ze elfmz jvez.

possiamo fare un'analisi di frequenza su questo messaggio, e lo mettiamo a **confronto** con la tabella di frequenza della lingua italiana

Messaggio cifrato

Lettera	Frequenza
V	12.12%
Z	11.52%
R	9.90%
F	9.09%
E	8.28%
I	7.88%
C	6.46%
K	5.45%
U	4.44%
J	4.44%
L	4.04%
T	3.64%
X	2.83%
G	2.42%
D	2.02%
M	1.82%
H	1.01%
W	1.01%
Y	0.81%
Q	0.61%
S	0.20%

Tabella di frequenza

Lettera	Frequenza
E	11.49%
A	10.85%
I	10.18%
O	9.97%
N	7.02%
T	6.97%
R	6.19%
L	5.70%
S	5.48%
C	4.30%
D	3.39%
U	3.16%
P	2.96%
M	2.87%
V	1.75%
G	1.65%
H	1.43%
B	1.05%
F	1.01%
Z	0.85%
Q	0.45%

Con questo capiamo che molto probabilmente una lettera tra **V**, **Z** oppure **R** nel testo cifrato corrisponderà alla **E** nel messaggio in chiaro. E Chiaramente questo ragionamento si può applicare a tutte le altre lettere, che con una buona probabilità assomiglierà ad una delle lettere con la percentuale simile.

N.B. usando l'alfabeto italiano le lettere straniere (**K**, **J**, **W**, **X** e **Y**) non sono presenti nella tabella di frequenza, ma con la cifratura che ho fatto io nel messaggio cifrato appaiono però mancheranno 5 lettere dell'alfabeto italiano perchè non hanno nessuna corrispondenza con le lettere straniere mancanti

In più si possono fare anche delle **osservazioni linguistiche**, per esempio in italiano solamente le lettere **A**, **E** e **O** possono stare da sole (per indicare le relative funzioni di preposizione), quindi nel messaggio cifrato possiamo sicuramente dire che le lettere **V** e **R** possono essere solamente **A**, **E** oppure **O**, che effettivamente combacia anche con le percentuali delle tabelle. Ad ogni modo, proviamo a sostituire nel messaggio cifrato le lettere con la percetuale più simile:

hder tivo ser ripo sa tovo, ufe gorpe i vezzopaotno, lti sde uilene non anlettolle sa vonla, ldllo i cena e i porba, i ceuonsi serro cmotpete e ser taenltite sa hderra, gaen, hdica i dn ltillo, i tacltanpetca, e i mtenset uotco e bapdti sa badve, lti dn mtovonlotao i seclti, e dn'ivmai uoclaeti sirr'irlti mitle; e ar monle, ufe aga uonpadnpe re sde tage, mit ufe tensi inuot mad cencaqare irr'ouufao hdecli lticbotvizaone, e cepna ar mdnlo an uda ar ripo uecci, e r'rssi tanuovanuai, met tamaprait moa nove sa ripo soge re tage, irronlininsoca sa ndogo, ricuain r'iuhdi saclensetca e tirrenlitca an ndoga porba e an ndoga cena. Ora come ora non sembra migliorato molto ma cominicamo a sistemare. Per esempio, come dicevamo prima le lettere singolo possono essere solamente **E**, **A** o **O**, ma ora abbiamo **I** e **E**, quindi dobbiamo cambiare la **I** e visto che è molto vicina di percentuale (nella tabella di frequenza) alla A, proviamo a invertire le due lettere. Così diventa

hder tavo ser rapo si tovo, ufe gorpe a vezzopiotno, lta sde ualene non inlettolle si vonli, ldllo a ceni e a porbi, a ceuonsa serro cmotpete e ser tienltate si hderri, gien, hdaci a dn ltallo, a ticltinpetci, e a mtenset uotco e bipdta si bidve, lta dn mtovonlotio a seclta, e dn'avmia uoclieta sarr'arlta matle; e ir monle, ufe igi uonpidnpe re sde tige, mat ufe tensa anuot mid cenciqire arr'ouufio hdecla ltacbotvazione, e cepni ir mdnlo in udi ir rapo uecca, e r'rssa tinuovinuia, met timipriat moi nove si rapo soge re tige, arronlanansoci si ndogo, racuian r'auhda siclensetci e tarrenlatci in ndogi porbi e in ndogi ceni.

Ora possiamo notare che ci sono molti **SI**, che molto probabilemte possono essere **DI**, quindi proviamo a invertire **D** e **S**

hser tavo der rapo di tovo, ufe gorpe a vezzopiotno, lta dse ualene non inlettolle di vonli, lslllo a ceni e a porbi, a ceuonda derro cmotpete e der tienltate di hserri, gien, hsaci a sn ltallo, a ticltinpetci, e a mtendet uotco e bipsta di bisve, lta sn mtovonlotio a declta, e sn'avmia uoclieta darr'arlta matle; e ir monle, ufe igi uonpisnpe re dse tige, mat ufe tenda anuot mis cenciqire arr'ouufio hsecla ltacbotvazione, e cepni ir msnlo in usi ir rapo uecca, e r'rdda tinuovinuia, met timipriat moi nove di rapo doge re tige, arronlanandoci di nsogo, racuian r'auhsa diclendetci e tarrenlatci in nsogi porbi e in nsogi ceni.

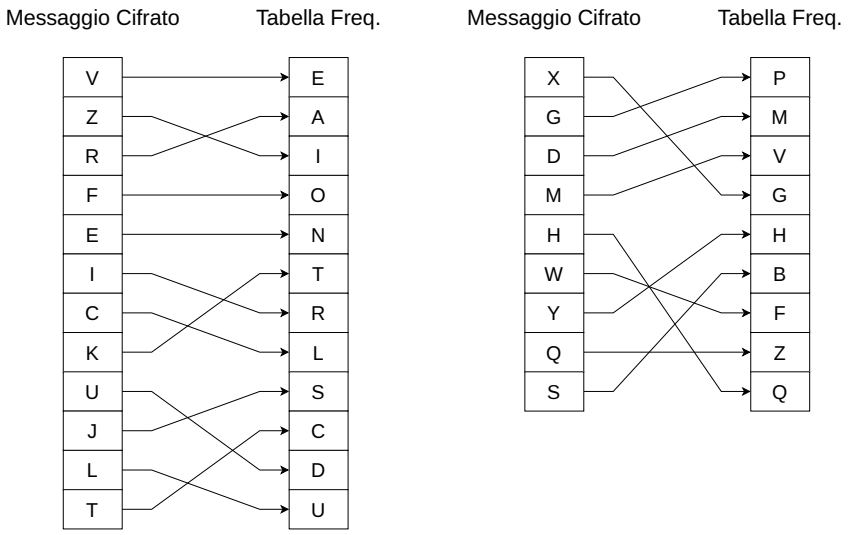
Ora siamo abbastanza sicuri che le lettere **D**,**A**,**I** e **E** siano al posto corretto. Possiamo continuare a fare delle anazilisi e man mano dedurre il testo originale. Per esempio nella prima riga c'è **DSE** e molto probabilmente potrebbe essere **DUE** (visto che la D ed E siamo sicuri che siano giuste). Oppure la parola **vezzopiotno** potrebbe essere **mezzogiorno** e così via. Se vuoi provare a risolvere te questo enigma da solo, sennò continua che ora c'è la soluzione.

Ad ogni modo, con un pò di pazzienza potreste vedere che il testo cifrato qua sopra non è altro che i primi versi dei **Promessi Sposi**

quel ramo del lago di como, che volge a mezzogiorno, tra due catene non interrotte di monti, tutto a seni e a golfi, a seconda dello sporgere e del rientrare di quelli, vien, quasi a un tratto, a ristringersi, e a prender corso e figura di fiume, tra un promontorio a destra, e un'ampia costiera dall'altra parte; e il ponte, che ivi congiunge le due rive, par che renda ancor piu sensibile all'occhio questa trasformazione, e segni il punto in cui il lago cessa, e l'adda rincomincia, per ripigliar poi nome di lago dove le rive, allontanandosi di nuovo, lascian l'acqua distendersi e rallentarsi in nuovi golfi e in nuovi seni.

N.B. Per semplicità ho messo tutto in minuscolo e ho rimosso le lettere accentate, ma chiaramente se contemplate possono essere di aiuto per scoprire parole all'interno del testo.

Ora vediamo le corrispondenze quanto era giuste



Si può notare che nel complesso questa tecnica quasi tutte le lettere le aveva quasi indovinate, infatti molte lettere erano sbagliate di una casella.

- Indovinate: $4/21 \approx 19.0$ %
- Sbagliato di una casella: $10/21 \approx 47.6$ %
- Sbagliato di due caselle: $5/21 \approx 23.8$ %
- sbagliato di tre caselle: $1/21 \approx 4.8$ %
- sbagliato di quattro caselle: $1/21 \approx 4.8$ %