CENNI STORICI SULLA CRITTOGRAFIA

Alunni: Francesco Magliocco, Maria Ludovica Naccarati, Gabriele Rubeo, a. s. 2013 – 2014, IV Liceo Scientifico, Istituto Santa Maria, Roma.

Referente: prof. Davide Passaro

*****

Questo articolo nasce a seguito di una conferenza sulla Crittografia tenuta dagli autori di questo testo all'interno di un ciclo di incontri sulla matematica che si sono svolti presso l'Istituto Galileo Galilei di Roma.

Per affrontare questo tema è utile partire dal significato stesso del nome. Il termine "crittografia" deriva dal greco *graphè* che vuol dire scrittura e *kriptos* che vuol dire nascosto; questa branca della matematica studia infatti tutte le procedure da utilizzare per rendere un messaggio incomprensibile a tutte quelle persone che non siano né il mittente né il destinatario.

In questo articolo si effettuerà un breve *excursus* storico sulla crittografia per vedere come essa abbia interessato gli uomini, già a partire dal V secolo a. C.

Uno dei più antichi cifrari utilizzati è stato quello di Cesare, un cifrario per sostituzione in cui ogni lettera viene spostata di un numero fisso di posti; Cesare ad esempio aveva deciso di sostituire ogni lettera con quella che la seguiva di tre posti nell'alfabeto. Un tale sistema monoalfabetico è però facilmente violabile: basta

conoscere le proprietà statistiche del linguaggio con cui il testo è stato scritto, in altre parole conoscere le frequenze con le quali le lettere dell'alfabeto compaiono in un generico testo scritto nella stessa lingua del messaggio cifrato, per arrivare a trovare la corrispondenza tra le lettere del testo cifrato e quelle dell'alfabeto in chiaro. A questo proposito sono stati creati da noi due programmi, uno che analizza le frequenze di ogni lettera nelle varie lingue e l'altro che è un simulatore del cifrario di Cesare.

Programmi realizzati

Il cifrario di Cesare, come si è detto, è inefficace per alcuni aspetti legati alla possibilità di ottenere le frequenze caratteristiche di ciascuna lettera di una lingua. In questo paragrafo si illustrano due programmi da noi realizzati che evidenziano questa problematica: il primo, realizzato con Python, è un estrattore di frequenze da un testo dato; il secondo, realizzato con Visual Basic, permette di cifrare e decifrare un testo tramite il cifrario di Cesare.

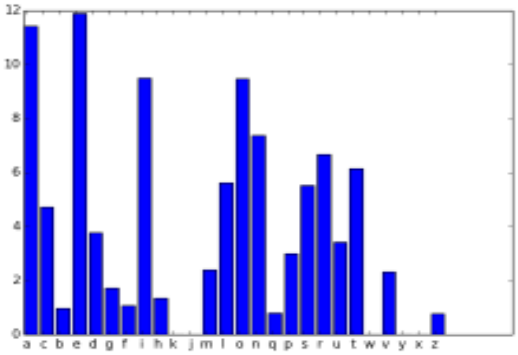
Python e Visual Basic sono due linguaggi di programmazione. In generale, un linguaggio di programmazione è un linguaggio formale dotato di un lessico, di una sintassi e di una semantica. È formato da due elementi: un compilatore, che traduce il programma sorgente, scritto in linguaggio di alto livello, più intuitivo e comprensibile dal programmatore, in un programma direttamente eseguibile dal calcolatore, scritto in linguaggio macchina o di basso livello; un interprete, che esegue il programma tradotto dal compilatore. Bisogna sottolineare che questa distinzione, valida in generale per molti linguaggi di programmazione, non riguarda né Python né Visual Basic, linguaggi direttamente interpretati.

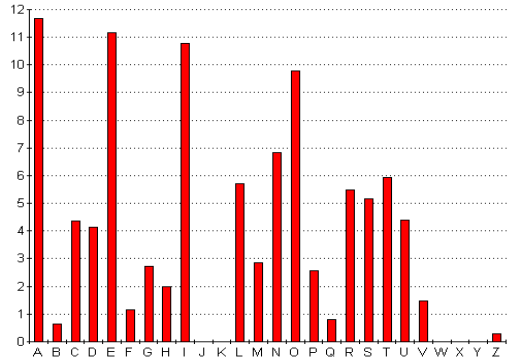
Si inserisce di seguito una descrizione dei due singoli programmi.

Estrattore frequenze

Questo programma permette di estrarre le frequenze con cui ciascuna lettera dell'alfabeto si presenta in un dato testo fornito in input. Il programma rende evidente l'inefficacia del cifrario di Cesare, dovuta al fatto che, in ciascuna lingua, ogni lettera in un alfabeto si presenta con una particolare frequenza: analizzando le frequenze si può capire con quali lettere, nel cifrario, le lettere dell'alfabeto sono state sostituite. All'interno del codice sviluppato, un dizionario fa corrispondere a ogni lettera dell'alfabeto la frequenza che questa ha nel testo, considerando insie-

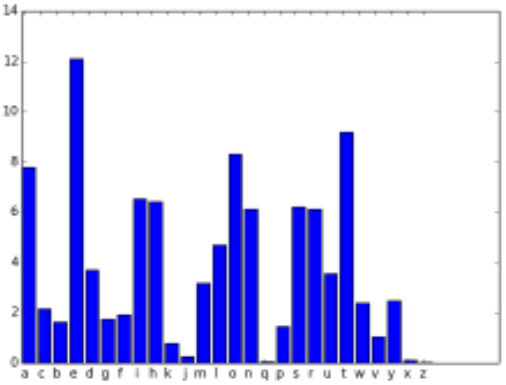
me le frequenze di una lettera e della corrispondente maiuscola; successivamente viene fatto l'istogramma di questo dizionario. Si riporta una analisi comparativa di due grafici delle frequenze della lingua italiana ottenuti uno con la lettura, attraverso il nostro programma, de "I promessi sposi" di Alessandro Manzoni, l'altro dal web:

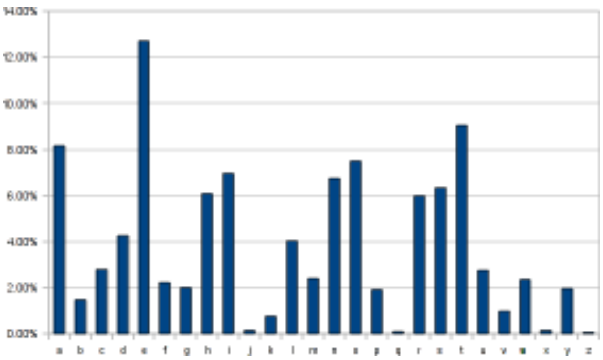




Le frequenze dei due istogrammi sono circa coincidenti.

Si sono analizzate anche le frequenze della lingua inglese, in particolare delle lettere del testo "Romeo and Juliet" di William Shakespeare e, in modo analogo a caso precedente, sono state confrontate con quelle ottenute dal web. Questo esempio mostra chiaramente che ogni lingua ha delle frequenze caratteristiche e che, ad esempio, quelle della lingua italiana sono differenti da quelle della lingua inglese.





Criptatore

Il secondo programma realizzato permette di criptare e decriptare allo stesso modo un testo dato facendo corrispondere, secondo diverse modalità, a ogni valore del codice ASCII, un sistema di codifica dei caratteri comune nei calcolatori, un altro suo valore. Visual Basic, rispetto a Python, ha una maggiore facilità di realizzazione di una interfaccia grafica che, nel caso specifico, si compone di due caselle di testo in cui si inseriscono, rispettivamente, i testi da cifrare e da decifrare e due in cui il programma risponde con il testo cifrato e decifrato; di due bottoni, per cifrare e decifrare; il programma offre inoltre la possibilità di cambiare modalità di cifrazione cambiando le corrispondenze tra testo in chiaro e testo cifrato nel codice ASCII. Nel codice, tramite un ciclo for, ogni carattere del testo viene sostituito con il carattere che corrisponde al valore ASCII del carattere originario diminuito di un certo valore.

Ad esempio, il testo *"Siamo studenti dell'Istituto Santa Maria."* viene cifrato in *“Nd\hj_nop_`iod__`gg Dnodopoj_N\io_H\md\)”*. La S con cui inizia il messaggio in chiaro corrisponde al valore 83 del codice ASCII, a cui viene sottratto il valore 5, che può essere scelto arbitrariamente. Il programma restituisce il valore letterale del codice ASCII corrispondente al valore numerico 83 - 5 = 78, ovvero "N". Questo processo avviene per tutti gli altri caratteri del messaggio.

Vigenere ed Enigma

Dopo aver scardinato il cifrario di Cesare, nel corso degli anni sono nati altri metodi di cifratura. Uno di questi è la Tavola di Vigenere. Il principale punto di forza di questo metodo è l’utilizzo non di uno ma di ben 26 alfabeti cifranti per cifrare un solo messaggio.

Questo metodo può essere visto come una generalizzazione del cifrario di Cesare; infatti invece di spostare sempre dello stesso numero di posti la lettera da cifrare, questa viene spostata di un numero di posti variabile, determinato dalle lettere della parola chiave concordata tra mittente e destinatario.

Il punto debole di questo cifrario è però che in un crittogramma di Vigenère si trovano spesso sequenze identiche di caratteri a una certa distanza l’una dall’altra. Se allora si individuano tutte le sequenze ripetute è certo che il massimo comun di-

visore tra le distanze tra sequenze identiche rappresenta la lunghezza della chiave.
Sicuramente scardinare un messaggio criptato con il metodo di Vigenere risulta più complicato rispetto allo scardinare un messaggio criptato con il cifrario di Cesare, ma anche questo metodo nel corso degli anni si è rivelato fallibile.

Scardinato il codice di Vigenère, inizia la ricerca di un nuovo codice di cifratura. Negli anni venti del '900 si arriva allora alla macchina di cifratura chiamata Enigma. Il principio su cui si basa questa macchina è quello di cambiare la cifratura per sostituzione ogni volta che la lettera viene codificata.
Questa macchina può essere immaginata come un grande cilindro costituito da tre tamburi. In cima al cilindro ci sono 26 fori disposti lungo il bordo, corrispondenti alle 26 lettere dell'alfabeto. Per codificare una lettera si lascia cadere una biglia nel foro corrispondente a quella lettera. La biglia cade nel primo tamburo, che ha 26 fori nel bordo della faccia superiore e 26 lungo il bordo della faccia inferiore, e attraversa dei tubi che collegano i fori superiori e quelli inferiori; questi tubi inoltre si piegano e si torcono di modo che una biglia che entra dalla sommità del tamburo emergerà dal fondo in una posizione completamente differente. Il tamburo centrale e quello inferiore sono simili, ma i tubi che li collegano uniscono i fori superiori a quelli inferiori in modi diversi; quando la biglia fuoriesce da uno dei fori sul fondo del terzo tamburo entra nell'ultima componente del "marchingegno" ed emerge da uno dei 26 fori sul fondo del cilindro, a ciascuno dei quali è associata una lettera dell'alfabeto.

Se il nostro "marchingegno" rimanesse nella stessa configurazione, non rappresenterebbe altro che un modo complicato per riprodurre una cifratura per sostituzione; la genialità di Enigma, invece, sta proprio nel fatto che ogni volta che una biglia cade attraverso il nostro cilindro il primo tamburo ruota di 1/26 di giro, perciò, quando la biglia successiva verrà fatta cadere nel cilindro, il primo tamburo le farà compiere un percorso completamente diverso.
Inoltre ogni giorno gli operatori di enigma potevano cambiare la disposizione dei dischi; ogni giorno infatti essi disponevano di un cifrario grazie al quale sapevano, al principio di ogni giornata, come andavano disposti i dischi per codificare i messaggi e il destinatario avrebbe decodificato il messaggio adottando la medesima regolazione tratta dal cifrario comune.

Con il tempo, però, i matematici compresero che ciascuna regolazione dei dischi possedeva alcune caratteristiche peculiari e che quelle regolarità potevano essere sfruttate per procedere a ritroso e decodificare i messaggi cifrati. Se l'operatore avesse digitato una A la lettera sarebbe stata cifrata, in base alla disposizione dei dischi, per esempio, in una D; a quel punto il primo disco sarebbe scattato di una posizione. Se, quando si digitava un'altra A essa veniva codificata con una Z, allora in un certo senso la A era collegata alla Z dal modo in cui i dischi erano stati configurati. Ogni giorno gli operatori tedeschi di Enigma usavano la stessa configurazione dei dischi, che trovavano nel cifrario. Quindi sceglievano la configurazione e per non commettere errori erano invitati a digitare due volte la configurazione scelta. Questa mossa si è rivelata quindi un indizio utile ai polacchi per comprendere come i dischi collegavano le lettere, quindi un'indicazione del modo in cui la macchina era configurata quel giorno.

Un gruppo di matematici inglesi ha studiato le regolarità individuate e ha realizzato il modo per automatizzare la ricerca delle configurazioni di Enigma tramite una macchina di loro stessa invenzione che hanno chiamato "Bomba".
Dopo aver scardinato anche la macchina di cifratura Enigma, si è passati alla crittografia moderna che oggi suscita l'interesse di molti matematici.

RSA

Concluso l'*excurus* storico, in questo paragrafo si affronterà la crittografia moderna. In questo ambito il sistema più importante e famoso è RSA.

Per spiegarlo è necessario analizzare due dei sistemi crittografici che hanno preceduto cronologicamente RSA e sui quali questo è basato: il sistema Diffie Hellmann Merkle e la crittografia a chiave pubblica.

Si inizierà spiegando il primo dei due attraverso un semplice esempio: ipotizziamo che il mittente (Alice) debba inviare un messaggio cifrato al destinatario (Bob). Supponiamo anche che il messaggio corrisponda a una valigetta e il codice di cifratura a un lucchetto. Alice invia la prima volta la valigetta a Bob proteggendola con un lucchetto di cui solo lei possiede la chiave; una volta che Bob riceve la valigetta non può aprirla perché non ha la chiave del lucchetto applicato da Alice, perciò lui appone un ulteriore lucchetto alla valigetta di cui solo lui ha la chiave, così rimanda la valigetta ad Alice che, una volta ricevutala, toglie il lucchetto che lei stessa aveva applicato e invia nuovamente la valigetta a Bob. A questo punto Bob

ottiene la valigetta con applicato solo il suo lucchetto, così può levarlo, possedendone la chiave, e aprire la valigetta, potendo in tal modo leggere il messaggio decodificato.

I due lucchetti corrispondono a due funzioni applicate in successione che devono poter essere invertite per la decodifica del messaggio.

Molto presto i matematici si accorsero, però, che questo sistema causava diversi problemi: il primo era che sia il mittente che il destinatario, per applicare i loro lucchetti, dovevano aspettare la risposta dell'altro e questo comportava un ritardo nello scambio delle informazioni; il secondo è che l'idea dell'inversione di due funzioni è un concetto non traducibile facilmente in un modello matematico. Infatti f(g(x)) è spesso diverso da g(f(x)).

Per far fronte a questi problemi nacque la crittografia a chiave pubblica, basata sull'uso di due chiavi, una pubblica, l'altra privata, generate in modo che sia impossibile ricavare l'una dall'altra. La prima serve per codificare i messaggi ed è di pubblico dominio, mentre la seconda per decodificare gli stessi ed è mantenuta segreta.

RSA nacque a seguito dei due sistemi appena descritti e ne risolse i principali problemi. Fu applicato per la prima volta nel 1978 da Rivest, Shamir e Adleman, dalle cui iniziali deriva il nome.

Cenni di aritmetica modulare

È necessaria, prima di affrontare il sistema RSA dal punto di vista matematico, una breve introduzione all'aritmetica modulare. L'aritmetica modulare si basa sul concetto di congruenze. Due numeri interi a e b si dicono congrui modulo n, con n naturale, se, divisi per n, danno lo stesso resto. In formula si scrive: $a \equiv b \bmod n$. Un altro concetto importante è quello di minimo residuo positivo: dati x ed a interi ed n naturale, con $0 \leq a \leq n-1$, se $x \equiv a \bmod n$, allora a si dice minimo residuo positivo di x modulo n e si indica con: $a = x \bmod n$. In altre parole, a è il resto della divisione tra x ed n. Per poter comprendere RSA è necessario descrivere due importanti scoperte fatte nel campo dell'aritmetica modulare, una di Fermat (il piccolo teorema), l'altra di Eulero. Definiamo un calcolatore a orologio: un calcolatore a orologio con un certo numero di ore è un calcolatore che restituisce il valore della divisione tra il valore ottenuto con un normale calcolatore e il numero delle ore. Fermat

scoprì che, dato un calcolatore a orologio con un numero primo p di ore, se si eleva un numero C alla p si ottiene lo stesso C, ovvero

$$C^p \bmod p = C$$

In generale, si può scrivere:

$$C^{k(p-1)+1} \bmod p = C$$

La scoperta che fece Eulero fu che su un calcolatore a orologio con un numero p x q di ore se si eleva un numero C alla k(p - 1)(q - 1) + 1 si ottiene la base C:

$$C^{k(p-1)(q-1)+1} \bmod (pxq) = C$$

Come funziona RSA?

In base alle precedenti premesse è possibile descrivere il funzionamento di RSA. Vi sono una chiave pubblica e una privata. La chiave pubblica è formata dalla copia (N, E), quella privata dalla coppia (N, D), con N = p x q e con p e q numeri primi molto grandi, vedremo poi perché. Un messaggio M viene cifrato in un messaggio C nel seguente modo:

$$C = M^E \bmod N$$

Si può notare che nella cifrazione è usata solo la chiave pubblica. Per decifrare C è necessario elevarlo ad un esponente D ottenuto con p e q - è spiegato avanti come - tale che:

$$C^D \bmod N = M$$

Dunque, applicando la scoperta di Eulero sopra descritta e svolgendo alcuni passaggi:

$$M^{ExD} \bmod N = M$$
$$M^{ExD} \bmod (p \times q) = M$$
$$EXD = k\,(p - 1)\,(q - 1) + 1$$
$$D = [k\,(p - 1)\,(q - 1) + 1] / E$$

dove k sarà scelto come il più piccolo intero tale che anche D sia intero.

Per decifrare C occorre avere D; per avere D occorre avere p e q, noti solamente a chi crea le chiavi pubblica e privata. L'unico modo che avrebbe una persona che non possedesse la chiave privata di ottenere D sarebbe quello di scomporre N nei fattori primi p e q, procedimento che, data la grandezza di questi, richiederebbe un tempo umanamente (e computazionalmente) molto lungo.

Il fascino e il mistero dei numeri primi

Nell'affrontare l'analisi dei moderni sistemi crittografici è emersa in tutta la sua forza l'importanza dei numeri primi. Il percorso d'approfondimento svolto sulla crittografia ha, infatti, portato direttamente a riflettere sulle proprietà dei numeri primi.

Questo paragrafo conclusivo vuole, quindi, offrire degli spunti di riflessione assolutamente non esaustivi su questi "numeri speciali".

I numeri primi, infatti, sono forse i numeri più affascinanti e misteriosi della matematica. Sembra quasi, quando se ne parla, che vi siano delle regolarità sfuggenti, non rigorose, ma probabilistiche. Ad esempio, i numeri primi sono collegati con la serie di Fibonacci, in cui ciascun termine è ottenuto, partendo dai termini 1, 1, sommando i due precedenti:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

Si può notare che ogni n-esimo termine, con n primo, è primo anch'esso. Tuttavia, non è sempre così: il 19-esimo termine, 4181, pur essendo 19 primo, non è primo.

Vi sono numerose congetture riguardo ai numeri primi, come la congettura di Goldbach, che afferma che ogni numero pari maggiore di 2 è esprimibile come somma di due numeri primi. Ad esempio, 4 = 2 + 2; 20 = 17 + 3. Questa congettura rimane ancora oggi indimostrata.

Un altro problema importante è quello dell'identificazione dei numeri primi. Si può usare come test il piccolo teorema di Fermat; tuttavia, questo fornisce un criterio necessario, ma non sufficiente, ovvero un criterio che dà la certezza, nel caso in cui non sia verificato, che il numero non è primo, ma non che lo sia, nel caso contrario. Il problema più grande relativo ai numeri primi è che non si conosce una formula che restituisca in ogni caso un numero primo, ma solo metodi di approssi-

mazione. Sono proprio alcune di queste caratteristiche dei numeri primi a costituire il fondamento dei sistemi crittografici moderni analizzati in questo testo.