Criptografía

ETSInf - UPV

Práctica 2

Criptoanálisis monoalfabético

---

1. Introducción

El objeto de esta práctica es el estudio de los sistemas de sustitución monoalfabética. En particular, se estudian los sistemas: Caesar, Afín y la sustitución monoalfabética general.

El trabajo a realizar por parte del alumno consiste en descifrar textos cifrados con los sistemas anteriores. Los textos estarán disponibles en Poliformat, y corresponden a textos castellanos que consideran exclusivamente los 27 símbolos en mayúsculas del alfabeto castellano. El objetivo es encontrar las claves de cifrado que permiten descifrar los criptogramas, o bien obtener los textos descifrados, bien en forma de fichero o impresos.

Los ficheros de datos tienen la extensión nb y contienen un texto cifrado cuya forma es una lista de los códigos ASCII de las letras que componen el criptograma. Si se elige trabajar con Mathematica, se pueden leer como sigue:

```
SetDirectory["directorio"];
fichero=OpenRead["nombre de fichero"];
InputStream[nombre de fichero];
c=Read[fichero];
Close[fichero];
```

Después de ejecutar estas instrucciones la variable c contiene la lista de códigos. Para evitar problemas con el alfabeto, éste se puede construir como sigue:

```
alfa="ABCDEFGHIJKLMNÑOPQRSTUVWXYZ";
alfabeto=Characters[alfa];
```

Para trabajar con la lista de letras hay que hacer lo que sigue:

```
c1=FromCharacterCode[c];
c2=Characters[c1];
```

La primera instrucción produce un string de las letras correspondientes a los códigos contenidos en la variable c. La segunda instrucción convierte el string en lista de Mathematica.

2. Criptoanálisis monoalfabético

El criptoanálisis de estos sistemas es sencillo y se basa en el análisis de frecuencias del texto cifrado y su comparación con las frecuencia del lenguaje castellano. Para evitar sesgos, se han calculado las frecuencias de aparición de los caracteres en el texto considerado para extraer los ficheros de prácticas. La siguiente tabla ordena los caracteres teniendo en cuenta la frecuencia calculada.

Frecuencia Alta		Frecuencia Media		Frecuencia Baja	
e	14.0	u	4.9	v	1.1
a	12.3	t	3.8	g	1.0
o	9.8	c	3.6	j	0.6
s	7.6	m	2.7	f	0.5
n	6.6	p	2.1	z	0.4
r	6.2	q	2.0	ñ	0.2
i	5.6	b	1.5	x	0.04
l	5.5	y	1.4	k	0.0004
d	5.3	h	1.2	w	0.0002

Para criptoanalizar un mensaje cifrado por desplazamiento es suficiente considerar que la letra más frecuente en el texto cifrado y la *'E'*. La distancia entre ambos símbolos nos dará con gran seguridad la clave utilizada.

En el caso del cifrado Afín basta considerar que el símbolo más frecuente corresponde a la *'E'* y el segundo más probable a la *'A'* (los dos caracteres más probables en castellano). Un sistema de ecuaciones nos permitirá calcular los parámetros del cifrado.

En el caso general de sustitución monoalfabética la ordenación de los símbolos teniendo en cuenta su frecuencia de aparición permite colocar con casi certeza las tres o cuatro primeras. El resto necesariamente debe obtenerse por tanteo viendo el sentido del texto. El orden de los bigramas y trigramas puede ayudar a encontrar algunas correspondencias más y reducir la búsqueda. A continuación se muestran los bigramas más frecuentes en español:

$$EN, ES, DE, OS, EL, AS, ER, RA, LA, RE, NT, UE, AD, ON, AL$$

y los primeros trigramas:

$$QUE, ENT, NTE, DEL, CON, IEN, LOS, ELA, ADE, EST, ODE, ION$$

Las frecuencias y la ordenación de k-gramas anteriores constituyen una estimación, suficiente para la tarea a abordar, pero una estimación. Por tanto conviene notar que es más fiable en el caso de las letras que el de los bigramas y mejor la de los bigramas que la de los trigramas.