Estructuras de Datos y Algoritmos

Práctica I - Curso 2012/13

Rompiendo el Código Enigma

Introducción y objetivos

Como un pequeño homenaje a Alan Turing en su año conmemorativo, las prácticas de este curso de Estructuras de Datos se van a basar en replicar los esfuerzos realizados por los aliados durante la segunda guerra mundial para descifrar el código Enigma, utilizado por el ejército alemán, tarea en la que Turing participó y tuvo un papel decisivo.

Para ello se van a proponer dos prácticas cuyo objetivo será encontrar la clave que se utilizo para cifrar un mensaje mediante el código o máquina enigma. En ambos casos el método será similar al usado por los aliados, el *cribado* o método de la "palabra probable": Se supondrá que el mensaje original contiene una determinada palabra y se buscarán las claves que pudieran generar su equivalente en el texto cifrado.

En esta primera práctica se utilizará un enfoque de **fuerza bruta**: Se descifrará el texto encriptado para todas las claves posibles (aprox. 1.000.000) y se filtrarán los mensajes descifrados en los que aparezca la **criba** o palabra probable.

En la segunda práctica se utilizará un **preprocesado** previo a la tarea de descifrado que permitirá, usando tablas de dispersión como estructura de datos básica, acortar considerablemente la tarea de descifrar mensajes individuales.

Descripción de la Máquina Enigma

Atención: En lo que sigue se va a proporcionar una descripción de la Máquina Enigma muy sucinta y basada únicamente en la información necesaria para la práctica. Si se desea disponer de más detalles existen muchos recursos en la web que pueden proporcionar una información más precisa, por ejemplo:

- http://en.wikipedia.org/wiki/Enigma_machine
- http://www.pbs.org/wgbh/nova/military/how-enigma-works.html
- http://enigmaco.de/enigma/enigma_es.html (Simulación web de una máquina enigma)

Para poder realizar las prácticas es preciso que primero dispongamos de un código que permita simular una Máquina Enigma en sus tareas de codificación de mensajes. En lo que sigue vamos a representar a una Máquina Enigma como una función que recibe un **texto** (el mensaje) y una **clave** compuesta por dos partes, la indicación de los 3 rotores de entre los 5 disponibles que se van a utilizar (y en que orden se disponen) y el estado inicial (ángulo de rotación) de cada uno de los rotores. Como resultado ésta función devuelve otro texto, el **mensaje codificado**.

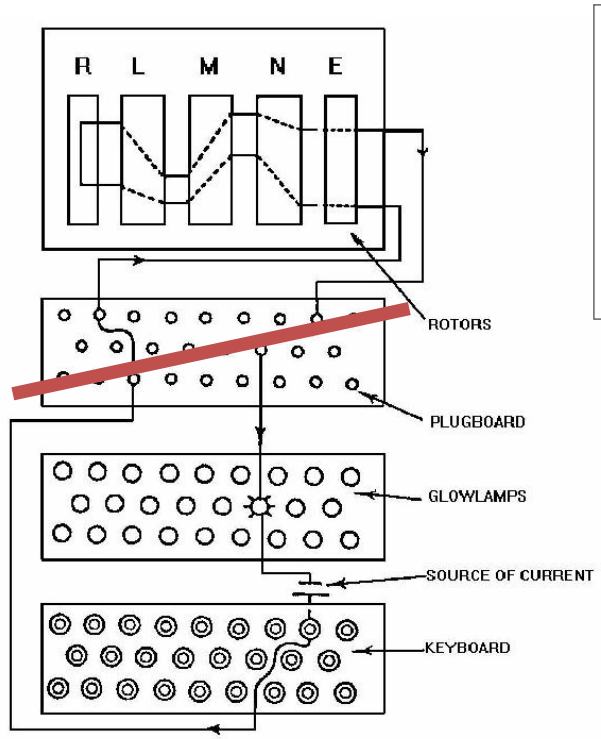
Por ejemplo, la máquina podría simularse por una función Java con la siguiente cabecera:

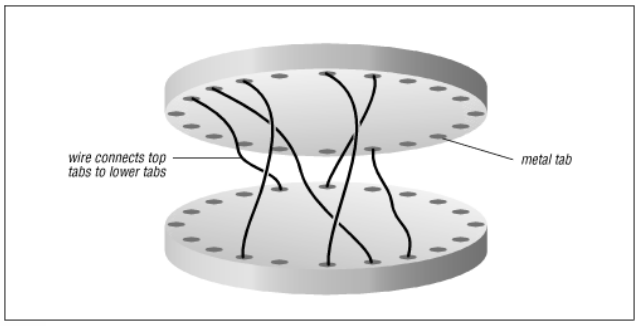
```
static String enigma(int[] r, String pos_ini, String txt)
```

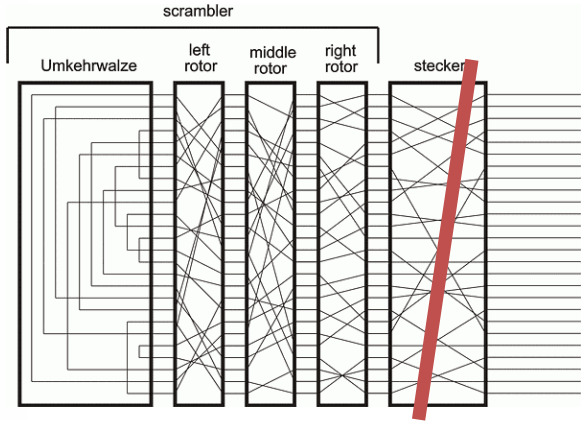
La Máquina Enigma tiene dos características fundamentales:

- En el texto únicamente pueden aparecer las 26 letras del alfabeto anglosajón, A-Z. No se permiten espacios en blanco (las palabras siguen unas a otras sin separación entre ellas), signos de puntuación ni números. Si se quiere incluir alguno de esos elementos en el texto se debe hacer usando una palabra o acrónimo que lo represente (PUNTO o CERO por ejemplo). A efectos prácticos esto significa que podemos (y debemos) representar un mensaje como una serie de números en el rango 0..25 (0-A, 1-B, 2-C, … 25-Z)
- La Máquina Enigma es **simétrica**: Cuando se cifra un mensaje con una determinada clave, si el texto cifrado se vuelve a cifrar con la máquina usando la misma clave se obtiene el mensaje original. Es decir, la clave de cifrado y descifrado es la misma.

Las partes principales de la Máquina Enigma son:







- Un teclado de 26 letras para introducir el texto original
- 3 rotores en secuencia, elegidos entre 5 rotores distintos disponibles.
  - Los rotores se disponen entre el teclado y el panel espejo, para nosotros el primero será el conectado directamente al teclado y el último el conectado al panel espejo.
  - Cada rotor es un cilindro con 26 contactos en cada cara, e internamente existen cables que conectan cada contacto de una cara con un contacto de la otra. Su objetivo es establecer una **permutación** fija entre sus 26 contactos (si entra corriente por un contacto de una cara se dirige a otro contacto predeterminado en su cara opuesta).
  - Cada rotor puede girar (26 posiciones antes de cada giro completo). Los contactos de un rotor siempre están alineados con los contactos del siguiente rotor.
  - Por último cada rotor tiene una pestaña asociada a un determinado contacto que provoca el avance en una posición del siguiente rotor cada giro completo (26 posiciones)
- Un panel espejo, que consiste en un disco circular con 26 contactos dispuestos únicamente en una cara (la conectada al último rotor). Cada contacto está conectado con otro.

- Un panel de 26 luces etiquetadas con letras que muestran el texto cifrado.

La forma de operar con una Máquina Enigma consiste en:

1. Se eligen 3 de los 5 rotores disponibles y se colocan en el orden adecuado (60 posibilidades distintas)
2. Se giran los 3 rotores a sus posiciones adecuadas. Cada rotor tiene 26 posiciones distintas, que están etiquetadas mediante letras. (26·26·26 = 17.576 posibilidades)
3. Los dos pasos anteriores sirven para establecer la **clave** con la que se va a cifrar (o descifrar) el mensaje. En total se tienen 1.054.560 posibles claves distintas.
4. Se pulsa una letra del teclado. Esto hace que avance una posición el primer rotor, y si su pestaña está en la posición adecuada (una vez cada 26 avances) a su vez provoca que avance el segundo rotor, y lo mismo con el tercero.
5. Cada letra tiene asociada un contacto del primer rotor, el cual desvía la corriente a otro contacto en su cara opuesta que entra en el segundo rotor, y así hasta llegar al panel espejo, el cual desvía la corriente a un contacto del tercer rotor y así en sentido inverso hasta salir por el primero, el cuál está asociado en la salida con una de las 26 lámparas la cuál se ilumina y muestra el carácter codificado.
6. Este proceso (pasos 4-5) se repite para cada letra sucesiva del mensaje. Normalmente se necesitaban dos personas para operar la máquina, una que pulsaba el teclado con el mensaje original y otra que apuntaba las letras de las lámparas que se iluminaban y que describían el mensaje cifrado.

Simulación de una Máquina Enigma

Para simular la Máquina Enigma lo primero que se debe tener en cuenta es que se va a traducir cualquier información de tipo textual (mensaje, los 3 caracteres que indican las posiciones iniciales de los rotores) a su equivalente numérico, traduciendo las letras adecuadamente (A-0, B-1, C-2, .. Z-25). Cuando se incrementen o decrementen estos valores que representan letras se utilizará **aritmética modular** para trabajar siempre en el rango 0..25 (por ejemplo, 20+10 = 4, 2-5 = 23)

Durante el proceso se van a utilizar los siguientes datos:

- **R**[1..3] → El vector que indica el número del rotor utilizado en cada posición (recordad que existen 5 rotores posibles que podemos usar). El rotor 1 es el más cercano al teclado y el 3 el que hace contacto con el panel espejo. Contiene valores en el rango 0..4, se proporciona como entrada y no cambia a lo largo del proceso.
- **P**[1..3] → El vector que contiene las posiciones (de rotación) actuales de cada rotor. Contiene valores en el rango 0..25. Se proporciona como entrada pero se va modificando a lo largo del proceso de codificación a medida que van girando los rotores.
- **TD**[0..4,0..25] → Contiene los 5 vectores (uno por cada distinto rotor) que representan el cableado interno (y la permutación que provocan) de cada rotor. Contienen valores en el rango 0..25. Ejemplo: **TD**[3,8] = 14 indica que el tercer rotor conecta su posición 8 (cara anterior) con la posición 14 (cara posterior) cuando la señal va en el sentido del teclado hacia el panel espejo.
- **TI**[0..4,0..25] → Contiene los 5 vectores (uno por cada distinto rotor) que representan el cableado interno (y la permutación que provocan) de cada rotor **cuando la señal va en dirección inversa** (del panel espejo hacia el teclado). Dado el ejemplo del punto anterior se tiene que cumplir que **TI**[3,14] = 8.

- **TE**[0..25] → El vector que indica el cableado del panel espejo. Contiene valores en el rango 0..25 y se cumple que, por ejemplo, si **TE**[3] = 12 entonces **TE**[12] = 3 (las posiciones 3 y 12 están conectadas).
- **M**[0..4] → El vector que indica la posición de las "pestañas" de cada rotor y permite saber cuando el avance de un rotor va a provocar el avance del siguiente.

Estos 4 últimos datos (**TD, TI, TE** y **M**) son conocidos (no se cambiaba el cableado interno de los rotores y podemos suponer que son "estándar").

	Rotor				
	0	**1**	**2**	**3**	**4**
Pestaña	16	4	21	9	25

	Rotor (TD)						Rotor (TI)				
Índice	**0**	**1**	**2**	**3**	**4**	**Espejo**	**0**	**1**	**2**	**3**	**4**
0	4	0	1	4	21	24	20	0	19	7	16
1	10	9	3	18	25	17	22	9	0	25	2
2	12	3	5	14	1	20	24	15	6	22	24
3	5	10	7	21	17	7	6	2	1	21	11
4	11	18	9	15	6	16	0	25	15	0	23
5	6	8	11	25	8	18	3	22	2	17	22
6	3	17	2	9	19	11	5	17	18	19	4
7	16	20	15	0	24	3	15	11	3	13	13
8	21	23	17	24	20	15	21	5	16	11	5
9	25	1	19	16	15	23	25	1	4	6	19
10	13	11	23	20	18	13	1	3	20	20	25
11	19	7	21	8	3	6	4	10	5	15	14
12	14	22	25	17	13	14	2	14	21	23	18
13	22	19	13	7	7	10	10	19	13	16	12
14	24	12	24	23	11	12	12	24	25	2	21
15	7	2	4	11	23	8	19	20	7	4	9
16	23	16	8	13	0	4	7	16	24	9	20
17	20	6	22	5	22	1	23	6	8	12	3
18	18	25	6	19	12	5	18	4	23	1	10
19	15	13	0	6	9	25	11	13	9	18	6
20	0	15	10	10	16	2	17	7	22	10	8
21	8	24	12	3	14	22	8	23	11	3	0
22	1	5	20	2	5	21	13	12	17	24	17
23	17	21	18	12	4	9	16	8	10	14	15
24	2	14	16	22	2	0	14	21	14	8	7
25	9	4	14	1	10	19	9	18	12	5	1

Dados estos datos, el algoritmo para simular la Máquina Enigma es el siguiente:

1. Traducir el texto (denominado **txt** en la cabecera de la función) a valores numéricos, lo representaremos como el vector **T**

2. Traducir la cadena de 3 caracteres que representa las posiciones iniciales de los rotores (denominada **pos_ini** en la cabecera de la función) a valores numéricos e inicializar el vector **P** con esos valores.
3. Recorrer todos los valores de **T**, y para cada uno de ellos, **T**[*i*], hacer lo siguiente:
4. **Girar los rotores**: Se incrementa **P**[1], pero previamente se comprueba si **P**[1] = **M**[**R**[1]] en cuyo caso se debe incrementar también **P**[2], pero previamente comprobando si hay que incrementar también **P**[3].
5. **Cifrado, recorrido directo**: Se comienza con el valor **X = T**[*i*] y se va transformando a su paso por los tres rotores. Si estamos en el rotor *k*-ésimo, se incrementa **X** con el valor **P**[*k*] (para obtener la posición relativa al rotor *k*), se permuta (**X ← TD**[**R**[*k*], **X]**) y se decrementa **X** con el valor **P**[*k*] (para volver a posición absoluta)
6. **Paso por el panel espejo**: Simplemente se permuta (**X ← TE**[**X**]) ya que éste panel no rota.
7. **Cifrado**, **recorrido inverso**: Igual que en el recorrido directo pero recorriendo los rotores en el orden 3,2,1.
8. El valor final de **X** corresponde al cifrado (o descifrado) de **T**[*i*].

Realización de la práctica

Tras haber codificado correctamente la función que simula la Máquina Enigma, la práctica consiste en realizar una aplicación (en Java o Python, si se desean otras alternativas consultar primero al profesor) que realice lo siguiente:

- Pida al usuario la palabra que va a servir de **criba**.
- Pida al usuario el texto cifrado que se intenta "romper".
- Entre en un bucle en el que recorra todas las posibles elecciones de rotores (3 enteros del 1 al 5 sin repeticiones – 60 posibilidades) y para cada una de ellas todas las posibles elecciones de posiciones iniciales (3 letras A..Z – 17.576 posibilidades) y para cada una de ellas codifique el texto cifrado (es decir, 1.054.560 descifrados del texto proporcionado). Cuando detecte la aparición de la palabra **criba** en un texto codificado lo escribirá por pantalla, junto con la clave (rotores y posiciones iniciales usados).
- Al final del proceso escribirá en pantalla el tiempo total empleado en el punto anterior.

Presentación y Evaluación de la práctica

Para una correcta evaluación de la práctica el alumno deberá:

1. Presentar electrónicamente, antes del 9 de noviembre de 2012, un fichero comprimido que contenga el código fuente de la aplicación utilizada para resolver el problema planteado.
2. Presentarse a la sesión de evaluación que le corresponda según su grupo de laboratorio el día 9 de noviembre de 2012. Es posible que en esa sesión se pida la modificación del código de la práctica y la obtención de nuevos resultados.

En el caso de realización por parejas (la situación habitual), tan sólo es necesario que uno cualquiera de ellos realice la presentación electrónica. En la evaluación, sin embargo, si es necesaria la presencia de ambos y la evaluación puede ser distinta para cada uno de ellos.

Caso de prueba

Si se introduce como palabra de **criba** el valor TURING y el mensaje cifrado es el siguiente:

```
RULPLIFAITTUEPTFBKMESPAQRPDIMLBHLUECOWHHHOKKBWBWQFJAPDXKD
JVBDEXVTWLTIHPBUGICXRGJVCHWHJHMHDIQFBLGWILJDQFXBSZZXUJIUI
HZGLMJBTGVAEIDZVJOIPBBFYJAFVVFPNWUEBIGWZTDBKYPHYNWOEAKWWM
NDXKGCQJUVIXYELIVHYZRQOYAKHKREAHUZCQMLKNEDZLJSWACOCWFNPEJ
VSFCBDBCIUMDJYRPAFONXYBIFBVDIHHNMXEGQAJLPQGEDHWFNTIBYRYQD
MSISXYSQQQBAWWGXNHVAELDPHOMUHFSCYUFPGGIBBRMASTRIFNHBRSAXQ
BLSJCVUEPMDDOWRSKSGAJNXBIQSWQAWEWLYWXILWUMFUNHHFRXGJNPNPG
TFTODFCDIVTCWVOISRJSFQRXCUBVVJFLRBRFFFIDEBLCOJCFWIUD
```

El programa deberá escribir por pantalla lo siguiente:

```
-------
432-FGC
XAPITWCZDBCAMKFXFREFPMJSUIPSIDHMATOOQXNUKXFQNZMDIWYFOOZHB
RKSVXLQBOPCSDKEXZEIGQORNOVIAYSCYEXIPREHOHAPXPKWJQLRWIRKGD
RIMCHEHBPYOITLMYNFDBIPPJANKOYKYTURINGNVPWKDCCFOCDLQHYVKMY
FFFHIIOIQLELJHXLNVTAGPCKCPFXFFJMGDFUTFUSBZGFBLPIIRSQJVWWU
LUQXIMYSMWZQXTACCQGZBFCGBJOVJKZOZFKEZTNQIHZGKLUAAGMJPVUVA
IUTKGVBUEJULTHWOUUXHYBIZTXSHNXTTHODRQPAEFGACQIHZSAQIJQZHK
FGROZLHVSAFXSKCYMIRRMCJSGAUJGCUAGWBGUHDXZYEKCCCVDYFVMCMHZ
MYQTFTUYHQSRQKALWBSRTSDHBJQXABYNTWSEZSWZPLWIZANCEYFT
-------
452-XTR
SISEDESEAQUEUNAMAQUINAIMITEELCOMPORTAMIENTODEUNACOMPUTADO
RAHUMANAENALGUNAOPERACIONCOMPLEJADEBEMOSPREGUNTARLEAESTAU
LTIMACOMOLOHACEYLUEGOTRADUCIRLARESPUESTAENLAFORMADEUNATAB
LADEINSTRUCCIONESPUNTOELDISENODEESTASTABLASUSUALMENTESEDE
NOMINAPROGRAMACIONPUNTOPROGRAMARUNAMAQUINAPARAQUEEFECTUEL
AOPERACIONCOMILLASACOMILLASSIGNIFICAINTRODUCIRENLAMAQUINA
LATABLADEINSTRUCCIONESAPROPIADAPARAQUEREALICECOMILLASACOM
ILLASPUNTOGUIONALANTURINGMILNOVECIENTOSCUARENTAYOCHO
-------
264.2 seg.
```