

## Práctica 1.2: Tablas Hash

**Ejercicio 1:** Implementar el método rehash de la clase `HashMap` cuya cabecera se muestra a continuación.

```
/** Doubles the size of the hash table and rehashes all the entries. */  
protected void rehash() {  
    // TO DO  
}
```

Además, se deberán definir pruebas unitarias asociadas a este método definiendo un caso de test Junit.

**Ejercicio 2:** La clase `HashMap` adjunta presenta un esquema de direccionamiento abierto basado en prueba lineal. Modificar e implementar las clases necesarias para dar también soporte a los esquemas basados en prueba cuadrática y hashing Doble. El código resultante deberá venir acompañado de su correspondiente diagrama de clases.

Nota: en la valoración del ejercicio se tendrá en cuenta la escalabilidad de la solución propuesta.

**Ejercicio 3:** Implementar una nueva clase llamada `HashMapSC` e introducir las modificaciones necesarias en la jerarquía de clases presentada en el ejercicio anterior, para poder trabajar también con tablas hash que utilicen un esquema de manejo de colisiones basado en encadenamiento separado (Separate Chaining). El código resultante deberá venir acompañado de su correspondiente diagrama de clases.

**Ejercicio 4:** Se desea desarrollar un programa que permita analizar las noticias web más relevantes del día. Para ello, todos los días se realiza un proceso de crawling para descargar las noticias más relevantes de un conjunto de webs preestablecidas. Estas noticias se postprocesan y se almacenan como ficheros de texto en un directorio. De esta forma el título de la noticia siempre se corresponderá con la primera línea del fichero, y el cuerpo con las líneas restantes.

En cualquier momento del día, se quiere poder extraer cierta información relevante de las noticias descargadas. En concreto sería interesante saber cuántas noticias podrían estar tratando un mismo tema. Para ello se desea obtener un listado en el que se especifiquen las distintas temáticas que tratan las noticias descargadas, así como los titulares de las noticias asociadas a una temática concreta.

Para conseguir esto, cada noticia debe analizarse por separado extrayendo cierta información asociada:

- Título de la noticia
- Entidades nombradas: una entidad nombrada es una cadena de texto que se refiere a nombres de personas, lugares u

organizaciones entre otros. Por cada documento deberán extraerse el conjunto de entidades nombradas que la conforman. Por entidad nombrada se entiende toda palabra que aparezca en el texto con el primer carácter en mayúscula y que no se encuentre dentro de la lista de palabras que se adjunta con el enunciado de la práctica (ES\_stoplist.txt).

- Frecuencia de cada una de las entidades nombradas que aparecen en la noticia.
- Entidad nombrada que aparece con mayor frecuencia en la noticia.

Una vez analizadas todas las noticias, la información extraída deberá mantenerse en memoria para poder realizar diferentes análisis. Puesto que la información extraída es utilizada por diferentes procesos analíticos que acceden de forma intensiva, se requiere que el acceso a la misma tenga lugar en el menor tiempo posible.

Por otro lado, se pide realizar un primer análisis de las noticias que consista en determinar qué noticias de las analizadas comparten una misma entidad nombrada como la más frecuente. Así, dado el conjunto de noticias almacenado en el directorio (news) se deberá obtener un listado como el que se muestra más abajo.

En él, se muestran los títulos de todas las noticias agrupados por la entidad que aparece con más frecuencia en el cuerpo de la noticia.

Puesto que al final del día se habrán descargado un elevado número de noticias, se pide que el proceso de extracción de información de cada una de ellas tenga lugar en el menor tiempo posible.

```
...
Ejército
* [Evacuan de Nahr a el Bared a las familias de los milicianos de Fatah a el Islam ]

FARC
* [Chávez se reunirá en Caracas con delegados de las FARC]
* [Chávez anuncia ante Álvaro Uribe que recibirá en Venezuela a un enviado de las FARC]
* [Las FARC rechazan el ofrecimiento de Chávez de liberar rehenes en Venezuela]

Islam
* [Líbano toma el control de Naher a el Bared ]

PJD
* [Los islamistas moderados acarician la victoria en las elecciones parlamentarias en Marruecos]
* [Los nacionalistas Istiqlal ganan las elecciones en Marruecos]

Fatah
* [El Ejército de el Líbano anuncia la muerte en combate de el líder de Fatah a el Islam]
* [a el Fatah refuerza la seguridad en el campo de refugiados palestinos de Ain el Helu]
```

Rhys

\* [Liberan a los dos sospechosos detenidos por el asesinato de un niño de 11 años en Liverpool]

...