

En este ejercicio vamos a analizar los mensajes de un grupo de Whatsapp. En concreto queremos obtener información sobre:

- Número de mensajes a lo largo del tiempo
- Número de mensajes según el día de la semana, o la hora del día
- Número de mensajes escritos por cada participante en el grupo
- Palabras más utilizadas en el grupo
- Palabras más características de cada participante en el grupo

Partimos de un archivo de log de un grupo de Whatsapp. El formato del fichero que genera Whatsapp varía según se trate de la versión Android o iOS.

- Android:
 - 26/02/16, 09:16 - Leonard: De acuerdo, ¿cuál es tu punto?
 - 26/02/16, 16:16 - Sheldon: No tiene sentido, solo creo que es una buena idea para una camiseta.
- iOS:
 - [26/2/16 09:16:25] Leonard: De acuerdo, ¿cuál es tu punto?
 - [26/2/16 16:16:54] Sheldon: No tiene sentido, solo creo que es una buena idea para una camiseta.

Como puedes observar, cada mensaje viene precedido por la fecha, la hora, y el nombre del usuario. El formato no es CSV, como en otros ejercicios, sino que es, digamos, más libre. Así, la fecha ocupa el principio de una línea, que puede venir seguida de una coma. Tras un espacio en blanco, viene la hora (que puede contener o no los segundos), seguida de un espacio, un guión y otro espacio, o bien de dos puntos y un espacio, según la versión de Whatsapp. Por último, tenemos el nombre del usuario, acabando en dos puntos, y tras otro espacio, aparece el texto del mensaje.

Con estos elementos queremos diseñar el tipo Mensaje con las propiedades:

- Fecha
- Hora
- Usuario
- Texto

También queremos diseñar el tipo Mensajes que contendrá los mensajes guardados en un fichero de log de WhatsApp y con la funcionalidad:

- Grupo de mensajes por usuario
- Grupo de mensajes por hora
- Grupo de mensajes por fecha
- Número de mensajes por usuario
- Número de mensajes por hora
- Número de mensajes por día del mes
- Número de mensajes por día de la semana
- Gráficos de barras con número de mensajes por hora, día del mes o día de la semana
- Palabras distintas en los mensajes
- Frecuencia de cada palabra

- Frecuencia de cada palabra para cada usuario
- Palabras características de un usuario. Un diccionario con la importancia de cada palabra usada por un usuario, que no sea una palabra hueca y se repita por encima
 - La importancia de una palabra es número proporcional al número de veces que el usuario ha usado esa palabra, y al mismo tiempo inversamente proporcional al número de veces que esa palabra ha sido usada por el resto de usuarios. Es decir para una palabra p y un usuario u :

$$i(u,p)=(f(u,p)/np(u))*(np(r)/f(r,p))$$

- Las palabras huecas vienen dadas en un fichero.