Ejercicios IA -Procesamiento de Lenguaje Natural
Aproximaciones empiricistas

1) Haz un clasificador de sentimiento de críticas de restaurantes a partir de 500 mensajes positivos y 600 negativos y la siguiente frecuencia de aparición de términos en los mensajes. En la tabla se muestra el número de mensajes positivos y negativos que contienen dicho término.

término	***mensajes positivos***	***mensajes negativos***
fabuloso	50	0
mejor	200	60
gustar	300	300
volver	200	200
caro	50	150
lamentable	0	30

Utiliza Naive Bayes para determinar el sentimiento de un mensaje que tenga los siguientes términos:

- mejor, gustar, volver, caro
- mejor, gustar, lamentable

2) Considera el siguiente corpus de frases para un sistema de escritura predictiva de un dispositivo móvil

<s> Yo voy bien </s>

<s> Voy a casa </s>

<s> Yo voy a comer a casa </s>

<s> Voy de cráneo </s>

<s> Voy a estudiar </s>

<s> Ya voy yo </s>

<s> Vamos a comer </s>

<s> Quiero comer </s>

- Utilizando bigramas y considerando las palabras tal y como aparecen (sin coger su raíz o su lexema) indica la siguiente palabra para las siguientes frases
  - <s> Ya voy
  - <s> Voy
- ¿Cómo cambia la predicción considerando trigramas?
- ¿Cuál es la frase que se generaría automáticamente utilizando bigramas?

3) Considera la siguiente tabla de frecuencias de aparición de términos en varios documentos.

	A	**B**	**C**	**D**	**E**
naranja	30	0	30	10	0
limón	0	10	10	0	10
kiwi	0	30	10	30	0
fresa	0	20	10	0	10
manzana	0	10	0	0	10
pera	10	0	0	0	0
piña	10	0	0	0	10

Calcula el ranking de documentos que recuperaríamos usando la similitud del coseno para la consulta "naranja" y usando como representación de los términos del documento lo siguiente:

- La presencia del término en el documento (variable binaria)
- La frecuencia del término en el documento
- El peso TF-IDF del término en el documento