

Ejercicios IA -Procesamiento de Lenguaje Natural

1) 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?

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

	A	B	C	D	E
naranja	10	40	0	10	0
limon	0	10	10	0	10
kiwi	0	10	0	10	0
fresa	30	0	0	10	0
manzana	10	0	40	10	0
pera	0	0	0	0	0
piña	0	0	0	0	0

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

3) 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.

<i>término</i>	<i>mensajes positivos</i>	<i>mensajes negativos</i>
<i>fabuloso</i>	50	0
<i>mejor</i>	200	60
<i>gustar</i>	300	300
<i>volver</i>	200	200
<i>caro</i>	50	150
<i>lamentable</i>	0	30

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

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