

Asignatura: MINERÍA DE TEXTOS 27/01/2021 11:30

- La nota del examen representa el 60% de la valoración final de la asignatura (el 40% restante corresponde a las prácticas).
- Cada cuestión contestada correctamente vale 1 punto.
- Se permite usar calculadora, pero ningún otro material ni realizar consultas por ningún medio.

Pregunta 1

Indique que son expresiones temporales absolutas, relativas y duraciones. Ponga tres ejemplos de cada una de ellas en inglés o en español.

Pregunta 2

Marque las expresiones temporales en el siguiente fragmento de texto:

Finally today, we learned that the space agency has finally taken a giant leap forward. Air Force Lieutenant Colonel Eileen Collins will be named commander of the Space Shuttle Columbia for a mission in December. Colonel Collins has been the co-pilot before, but this time she's the boss.

Pregunta 3

Considere el siguiente etiquetado gramatical (POS tags) de una oración, donde la segunda columna corresponde al gold standard y la tercera a las etiquetas asignadas por un etiquetador que se está evaluando. Construya la matriz de confusión (valores absolutos, no porcentajes) correspondiente incluyendo todas las etiquetas que aparecen en el etiquetado en el siguiente orden: DT IN JJ NN NNP PRP\$ VBD VBP. Reflexione sobre lo qué indica la matriz construida sobre el etiquetado analizado:

Japan NNP NNP

began VBD VBD

the DT DT

defence NN NN

of IN IN

their PRP\$ PRP\$

Asian JJ JJ

Cup NNP NN

title NN NN

with IN IN

a DT DT

lucky JJ JJ

2-1 JJ JJ

win VBP NN

against IN IN

Syria NNP NNP

in IN IN

a DT DT

Group NNP NNP

C NNP NNP

championship NN NN

match NN NN

on IN IN

Friday NNP NNP

. . .

Pregunta 4

Representa los siguientes documentos mediante bolsa de palabras en una matriz término-documento usando como función de pesado TF-IDF. Utiliza todas las palabras de cada documento.

Doc1: El perro se escapó de la casa y llegó hasta el banco del pueblo.

Doc2: La casa parecía estar a punto de caerse.

Doc3: ¡Vaya día más perro!

Doc4: Me quedé sentada en el banco esperando que pasara el tiempo.

A la vista de la matriz resultante, reflexiona sobre las características de este modelo de representación y función de pesado en cuanto a palabras carentes de significado relevante, posibles palabras polisémicas, y cualquier otro aspecto que quieras destacar.

Pregunta 5

A partir de la matriz término-documento de la pregunta anterior calcula la matriz de similitud o la de distancias (a elegir por el estudiante) entre documentos utilizando una medida de similitud o distancia entre vectores. Indica claramente qué medida has utilizado para el cálculo.

Pregunta 6

Supongamos que se pretende desarrollar una aplicación para un buscador web que al realizar una consulta con un nombre de persona nos devuelva los posibles diferentes individuos que comparten ese mismo nombre, junto con las páginas que corresponderían a cada uno de los individuos. ¿Cómo enfocarías la resolución de este problema, como un problema de clustering o como uno de clasificación automática? Justifica tu respuesta y describe una posible solución.