

Curso de IA

Capítulo 7. Cuestionario

Para estudiantes

© 2023 SAMSUNG. Todos los derechos reservados.

La Oficina de Ciudadanía Corporativa de Samsung Electronics posee los derechos de autor de este documento.

Este documento es una propiedad literaria protegida por la ley de derechos de autor, por lo que está prohibida su reimpresión y reproducción sin permiso.

Para utilizar este documento fuera del plan de estudios de Samsung Innovation Campus, debe recibir el consentimiento por escrito del titular de los derechos de autor..

1. ¿Cuál de las siguientes declaraciones sobre la minería de textos es incorrecta?

- 1 Se puede conocer la reacción a un grupo específico.
- 2 Se puede obtener información sobre el liderazgo de ventas.
- 3 Se pueden establecer estrategias competitivas monitoreando otras marcas.
- 4 Se puede aplicar el mismo método a varios idiomas.

2. ¿Cuál de las siguientes afirmaciones no es cierta sobre el uso de las redes sociales?

- ① Es posible conocer cuántos grupos componen la red.
- ② Es posible conocer clientes influyentes.
- ③ Es posible saber ver el cambio a lo largo del tiempo.
- ④ Es posible saber si los clientes se irán la próxima vez.

3. ¿Qué término significa separar una raíz de una palabra cuya forma ha sido modificada para extraer la palabra que es objeto del análisis morféxico?

R= lematización

4. ¿Qué término designa un conjunto de materiales que pueden mostrar los aspectos esenciales de la lengua como material de investigación requerido en cada campo de la investigación lingüística?

R= Corpus

5. Explique qué es la TF (Term Frequency) y escriba cómo calcular la TF (fórmula).

La TF, nos dice cuántas veces se repite una palabra en un documento

Se calcula así: $TF = \frac{\text{Número de veces que el término aparece en el documento}}{\text{Número total de términos en el documento}}$, es decir, $TF = t/d$, donde "t" es el termino específico y "d" el total de términos en el documento

6. Encuentre la declaración más similar utilizando TF-IDF y similitud coseno con referencia al siguiente código.

```
importar nltk
importar numpy como np
desde nltk.corpus importar stopwords
desde sklearn.feature_extraction.text importar TfidfVectorizer
desde sklearn.metrics importar pairwise_distances

doc = [
    "I es un campo en rápido avance que implica el desarrollo de máquinas inteligentes",
    "El aprendizaje automático es un subconjunto de la IA que se centra en la capacitación de máquinas para aprender de los datos.",
    " El aprendizaje profundo es un subcampo del aprendizaje automático que utiliza redes neuronales con múltiples capas",
    " La IA se aplica en diversos sectores, como la sanidad, el transporte y el ocio.
",
    " Las consideraciones éticas desempeñan un papel importante en el desarrollo de la IA",
    " La inteligencia artificial puede revolucionar muchos aspectos de la sociedad",
    " Los sistemas de IA, como los chatbots y los asistentes virtuales, son cada vez más comunes",
    "El procesamiento del lenguaje natural es una rama de la IA que permite a las máquinas comprender el lenguaje humano",
    " El objetivo de la visión por computadora es que las máquinas comprendan la información visual",
]

# preprocesamiento.
doc = [x.lower() for x in doc]

# parámetros
max_features = 18
min_df = 1
max_df = 3
stop_words = stopwords.words('english')

vectorizer = TfidfVectorizer(max_features=max_features,
                             min_df=min_df,
                             max_df=max_df,
                             stop_words=stop_words)
X = vectorizer.fit_transform(doc).toarray()
```

```

import nltk
import numpy as np
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import pairwise_distances
from sklearn.metrics.pairwise import cosine_similarity

```

✓ 0.0s

Python

```

doc = [
    "I es un campo en rápido avance que implica el desarrollo de máquinas inteligentes",
    "El aprendizaje automático es un subconjunto de la IA que se centra en la capacitación de máquinas para aprender de los datos.",
    "El aprendizaje profundo es un subcampo del aprendizaje automático que utiliza redes neuronales con múltiples capas",
    "La IA se aplica en diversos sectores, como la sanidad, el transporte y el ocio.",
    "Las consideraciones éticas desempeñan un papel importante en el desarrollo de la IA",
    "La inteligencia artificial puede revolucionar muchos aspectos de la sociedad",
    "Los sistemas de IA, como los chatbots y los asistentes virtuales, son cada vez más comunes",
    "El procesamiento del lenguaje natural es una rama de la IA que permite a las máquinas comprender el lenguaje humano",
    "El objetivo de la visión por computadora es que las máquinas comprendan la información visual",
]

# preprocesamiento.
doc = [x.lower() for x in doc]

# parámetros
max_features = 18
min_df = 1
max_df = 3
stop_words = stopwords.words('spanish')

```

```

vectorizer = TfidfVectorizer(max_features=max_features,
                             min_df=min_df,
                             max_df=max_df,
                             stop_words=stop_words)
X = vectorizer.fit_transform(doc).toarray()

```

✓ 0.0s

```
cosine_sim = cosine_similarity(X)
```

✓ 0.0s

☰ ▶ ⏪ ⏩ ☰ ...

```

np.fill_diagonal(cosine_sim, 0) # Evitar comparar un documento consigo mismo
most_similar = np.unravel_index(np.argmax(cosine_sim), cosine_sim.shape)

# Imprimir los documentos más similares
print("Los documentos más similares son:")
print("Documento 1:", doc[most_similar[0]])
print("Similitud coseno:", cosine_sim[most_similar])

```

✓ 0.0s

Los documentos más similares son:
 Documento 1: el aprendizaje automático es un subconjunto de la ia que se centra en la capacitación de máquinas para aprender de los datos.
 Similitud coseno: 0.6513390092841729