

Objetivo

Explorar y comparar las representaciones vectoriales de palabras para entender las similitudes y diferencias en cómo se representa las relaciones semánticas y sintácticas del lenguaje.

Descripción de la Práctica

1. Preparación del Entorno:

- Instalar Gensim y otras librerías necesarias.
- Descargar modelos pre-entrenados de Word2Vec

2. Toma de contacto:

- Familiarizarse con la API de Gensim para el acceso a los vectores de palabras.

3. Análisis Exploratorio:

- Seleccionar un conjunto de palabras para análisis (por ejemplo, "ciencia", "arte", "tecnología").
- Obtener los vectores correspondientes .
- Calcular y comparar la similaridad entre pares de palabras.

4. Tareas Específicas:

- Analogías: Resolver analogías como "París es a Francia lo que Tokio es a ¿qué?"
- Palabras más similares: Identificar las palabras más similares a un término dado en cada modelo.
- Divergencias: Encontrar ejemplos de palabras cuyas representaciones son notablemente diferentes.

5. Visualización:

- Utilizar técnicas de reducción de dimensionalidad (como PCA o t-SNE) para visualizar los vectores de un conjunto seleccionado de palabras.

```
from sklearn.manifold import TSNE

import matplotlib.pyplot as plt

# Lista de palabras a visualizar

palabras = ['rey', 'reina', 'hombre', 'mujer', 'príncipe', 'princesa']

vectores = [modelo[palabra] for palabra in palabras]

# Reducción de dimensionalidad

tsne = TSNE(n_components=2)

vectores_2d = tsne.fit_transform(vectores)

# Gráfica

plt.figure(figsize=(6,6))

for punto, palabra in zip(vectores_2d, palabras):

    plt.scatter(punto[0], punto[1])

    plt.text(punto[0] + 0.02, punto[1] + 0.02, palabra)

plt.show()
```