

CHATBOT DE VIDEOJUEGOS

Analisis de Similitud de Texto con TF-IDF

Carlos Navarro Martinez

Fecha: 29 de January de 2026

Proyecto: Chatbot sobre LOL, Valorant y CS:GO

Tecnologias: Python, sklearn, NLTK

1. INTRODUCCION

Este informe presenta la implementacion de un chatbot basico especializado en videojuegos competitivos (*League of Legends*, *Valorant* y *CS:GO*). El chatbot utiliza tecnicas de procesamiento de lenguaje natural (NLP) para responder preguntas de los usuarios mediante analisis de similitud de texto con TF-IDF y similitud del coseno.

Objetivo: Crear un chatbot capaz de comparar preguntas del usuario con una base de conocimiento predefinida y devolver respuestas relevantes cuando la similitud sea mayor o igual al 75%.

2. TECNOLOGIAS UTILIZADAS

Tecnologia	Descripcion
Python 3.x	Lenguaje de programacion principal
scikit-learn	TfidfVectorizer y cosine_similarity
NumPy	Operaciones numericas y arrays
NLTK	Procesamiento de lenguaje natural

3. METODOLOGIA

3.1. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF es una tecnica que convierte texto en vectores numericos considerando la importancia de las palabras en el documento y en todo el corpus. Se calcula como:
 $TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t)$

3.2. Similitud del Coseno

La similitud del coseno mide el angulo entre dos vectores. Valores cercanos a 1.0 indican alta similitud (vectores identicos), mientras que 0.0 indica sin similitud. En este chatbot se usa un **umbral de 0.75 (75%)** para determinar si dos preguntas son suficientemente similares.

4. BASE DE CONOCIMIENTO

El chatbot cuenta con **25 preguntas predefinidas** distribuidas en tres categorias:

Categoria	Cantidad	Temas
League of Legends	8 preguntas	Definicion, campeones, roles, Baron Nashor, farmeo, pentakills, runas
Valorant	9 preguntas	Definicion, agentes, rondas, Spike, armas, aces, roles, rangos, creditos
CS:GO	8 preguntas	Definicion, jugadores, AWP, rondas, skins, mapas, clutches, economia

5. RESULTADOS DE PRUEBAS

Se ejecutaron 18 preguntas de prueba automatizadas para verificar el funcionamiento del chatbot:

Metrica	Valor
Total de preguntas probadas	18
Respuestas encontradas ($\geq 75\%$)	4 (22.2%)
No entendidas ($< 75\%$)	14 (77.8%)
Similitud maxima	100.00%
Similitud minima	0.00%
Similitud promedio	56.18%

6. EJEMPLOS DE INTERACCION

6.1. Preguntas con Alta Similitud ($\geq 75\%$)

Pregunta: Que es Valorant?

Respuesta: Valorant es un shooter tactico en primera persona desarrollado por Riot Games...

Similitud: 100.00%

Pregunta: Que significa hacer un pentakill?

Respuesta: Un pentakill es cuando un jugador elimina a los cinco enemigos...

Similitud: 84.88%

6.2. Preguntas con Baja Similitud ($< 75\%$)

Pregunta: Puedes enseñarme a cocinar?

Respuesta: [X] No entiendo tu pregunta, intenta reformularla...

Similitud: 0.00%

7. INSTRUCCIONES DE USO

Instalacion de dependencias:

```
python -m pip install -r requirements.txt
```

Ejecucion del chatbot:

```
python chatbot.py
```

Ejecutar pruebas automatizadas:

```
python test_chatbot.py
```