

CHATBOT DE FÓRMULA 1

Chatbot con Reglas y Respuestas Basadas en Similaridad

Autor:	Julio García
Asignatura:	MIA - Modelos de Inteligencia Artificial
Tema:	Examen Tema 3
Fecha:	29 de enero de 2026

1. INTRODUCCIÓN

Este proyecto implementa un chatbot inteligente especializado en Fórmula 1, con enfoque particular en los pilotos españoles Fernando Alonso y Carlos Sainz Jr. El chatbot utiliza técnicas avanzadas de procesamiento de lenguaje natural (NLP) para comparar las preguntas del usuario con una base de conocimiento predefinida y responder de manera inteligente mediante análisis de similitud de texto.

Tecnologías Utilizadas:

- Python 3.x: Lenguaje de programación principal
- NLTK: Natural Language Toolkit para procesamiento de texto
- scikit-learn: Específicamente TfidfVectorizer para vectorización de texto
- NumPy: Para operaciones numéricas con vectores

2. EXPLICACIÓN DEL CÓDIGO

El chatbot se basa en un sistema de similitud de texto que compara la pregunta del usuario con un conjunto de preguntas predefinidas. A continuación se explican los componentes principales:

2.1. Vectorización TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) es una técnica que convierte texto en vectores numéricos. Asigna pesos a las palabras según su importancia en el documento:

- TF (Term Frequency): Frecuencia de una palabra en el documento
- IDF (Inverse Document Frequency): Importancia de la palabra en todo el corpus

El vectorizador utiliza los siguientes parámetros:

- lowercase=True: Convierte todo a minúsculas para normalizar
- ngram_range=(1,2): Captura palabras individuales y pares de palabras
- token_pattern: Patrón para identificar palabras válidas

2.2. Similitud del Coseno

La **similitud del coseno** mide el ángulo entre dos vectores en un espacio multidimensional. Los valores van de 0 a 1:

- 0: Vectores completamente diferentes (ángulo de 90°)
- 1: Vectores idénticos (ángulo de 0°)

El chatbot utiliza un **umbral del 30%** (0.3) para determinar si una respuesta es suficientemente relevante. Si la similitud es menor al 30%, el chatbot indica que no entiende la pregunta.

2.3. Flujo del Algoritmo

1. El usuario introduce una pregunta
2. La pregunta se vectoriza usando TF-IDF
3. Se calcula la similitud del coseno con todas las preguntas predefinidas
4. Se selecciona la pregunta con mayor similitud
5. Si la similitud $\geq 30\%$, se devuelve la respuesta correspondiente
6. Si la similitud $< 30\%$, se indica que no se entiende la pregunta
7. El proceso se repite hasta que el usuario escriba "salir"

3. PREGUNTAS Y RESPUESTAS DISPONIBLES

El chatbot cuenta con 15 pares de preguntas y respuestas predefinidas, organizadas en cuatro categorías principales:

3.1. Fernando Alonso

Pregunta	Respuesta
¿Quién es Fernando Alonso?	Fernando Alonso es un piloto español de Fórmula 1, considerado uno de los mejores de la historia. Nació en Oviedo en 1981 y es bicampeón mundial de F1.
¿Cuántos campeonatos mundiales ha ganado Fernando Alonso?	Fernando Alonso ha ganado 2 campeonatos mundiales de Fórmula 1, en 2005 y 2006 con el equipo Renault.
¿En qué equipo corre Fernando Alonso actualmente?	Fernando Alonso corre actualmente para el equipo Aston Martin desde la temporada 2023.
¿Cuál es la nacionalidad de Fernando Alonso?	Fernando Alonso es español, nacido en Oviedo, Asturias.

3.2. Carlos Sainz Jr

Pregunta	Respuesta
¿Quién es Carlos Sainz Jr?	Carlos Sainz Jr es un piloto español de Fórmula 1, hijo del legendario piloto de rallies Carlos Sainz. Nació en Madrid en 1994.
¿En qué equipo corre Carlos Sainz Jr actualmente?	Carlos Sainz Jr corre actualmente para el equipo Ferrari desde la temporada 2021.
¿Cuántas victorias tiene Carlos Sainz Jr en F1?	Carlos Sainz Jr ha conseguido 3 victorias en Fórmula 1: Gran Bretaña 2022, Singapur 2023 y Australia 2024.

3.3. Fórmula 1 General

Pregunta	Respuesta
¿Qué es la Fórmula 1?	La Fórmula 1 es la máxima categoría del automovilismo mundial, donde compiten los mejores pilotos en monoplazas de alta tecnología.
¿Cuántos equipos hay en la Fórmula 1?	En la Fórmula 1 hay 10 equipos, cada uno con 2 pilotos, sumando un total de 20 pilotos en la parrilla.
¿Cuál es el circuito más rápido de F1?	El circuito más rápido de F1 es Monza en Italia, conocido como el "Templo de la Velocidad", donde se alcanzan velocidades superiores a 350 km/h.
¿Quién tiene más campeonatos mundiales en F1?	Lewis Hamilton y Michael Schumacher comparten el récord con 7 campeonatos mundiales cada uno.
¿Cuál es el equipo más exitoso de la historia de F1?	Ferrari es el equipo más exitoso de la historia de F1, con 16 campeonatos de constructores y múltiples títulos de pilotos.
¿Qué significa DRS en Fórmula 1?	DRS significa "Drag Reduction System" (Sistema de Reducción de Resistencia). Es un alerón trasero móvil que reduce la resistencia aerodinámica para facilitar los adelantamientos.
¿Cuántos puntos se dan por ganar una carrera?	Por ganar una carrera se otorgan 25 puntos al piloto ganador. El segundo lugar recibe 18 puntos y el tercero 15 puntos.

3.4. Relación entre Pilotos

Pregunta	Respuesta
¿Qué relación tienen Fernando Alonso y Carlos Sainz?	Fernando Alonso y Carlos Sainz Jr son ambos pilotos españoles de F1. Alonso es una inspiración para Sainz, quien lo considera un referente. Ambos han representado con orgullo a España en la máxima categoría.

4. CAPTURAS DE PANTALLA

A continuación se muestran capturas de pantalla del chatbot en funcionamiento, demostrando interacciones reales con el sistema:

Captura 1: Interfaz del Chatbot

```
Tu pregunta: cual es el equipo con mas exito de la formula 1?  
Respuesta: Ferrari es el equipo más exitoso de la historia de F1, con 16 campeonatos de constructores y múltiples títulos de pilotos.
```

Captura 2: Ejemplo de Pregunta y Respuesta

```
PS C:\Users\jgortiz\Desktop\IA-BigData> & "C:/python 3.10.0/python.exe" "c:/Users  
=====  
CHATBOT DE FORMULA 1 - Fernando Alonso & Carlos Sainz Jr  
=====  
  
Hola! Soy un chatbot especializado en Formula 1.  
Puedo responder preguntas sobre:  
- Formula 1 en general  
- Fernando Alonso  
- Carlos Sainz Jr  
  
Escribe 'salir' para terminar la conversacion.  
=====  
  
Tu pregunta:
```

5. CONCLUSIONES

Logros Alcanzados:

- Implementación exitosa de chatbot basado en similitud de texto
- 15 pares de preguntas y respuestas sobre F1, Fernando Alonso y Carlos Sainz Jr
- Sistema de umbral funcional (30%) para determinar respuestas válidas
- Manejo de múltiples preguntas hasta comando "salir"
- Generación automática de informe de interacciones
- Código completamente documentado con comentarios explicativos

Aprendizajes:

- La similitud del coseno es efectiva para comparar textos cortos
- El umbral debe ajustarse según la variabilidad del lenguaje natural
- Los n-gramas mejoran significativamente la captura de contexto
- Una base de conocimiento amplia mejora la cobertura de respuestas

Posibles Mejoras Futuras:

- Implementar procesamiento de sinónimos para mayor flexibilidad
- Agregar más preguntas a la base de conocimiento
- Implementar aprendizaje incremental basado en interacciones
- Mejorar el manejo de acentos y caracteres especiales
- Agregar contexto conversacional (memoria de preguntas anteriores)