# Forma manual de capturar preguntas

Hay que definir criterios claros que permitan distinguir preguntas de otros mensajes (que sean respuestas, saludos, memes, spam, enlaces, entre otros). Algunas estrategias podrían ser:

* Detectar uso de signo de interrogación
* Identificar frases o palabras interrogativas
* Filtrar mensajes irrelevantes
* Revisar contexto del mensaje

También puede haber preguntas incluidas en respuestas y repreguntas.

## Detectar el uso de signos de interrogación

Se puede verificar si el mensaje empieza o termina con signos de interrogación. Sin embargo, esto no es suficiente porque algunos usuarios pueden hacer preguntas sin signo o usarlos en contextos que no son preguntas.

## Identificar patrones lingüísticos de preguntas

Las preguntas pueden contener:

* palabras interrogativas (qué, cómo, cuándo, dónde, por qué, quién, cuál, cuánto)
* frases comunes de consulta
  + Alguien sabe ...
  + Cómo se hace …
  + Me pueden ayudar con ...

Se puede armar una lista con estas palabras o frases comunes en preguntas y hacer una detección.

## Filtrar mensajes irrelevantes

Si el mensaje es solo un emoji, es probable que no sea una pregunta. También, hay que ignorar mensajes que solo tengan saludos y agradecimientos.

## Revisar el contexto del mensaje

Si el mensaje se encuentra presente en un hilo y ese id de hilo no está presente en la base de datos relacional, significa que puede ser una nueva pregunta.

## Preguntas incluidas en respuestas o repreguntas

Si la respuesta contiene signos de pregunta (?), se puede considerar como una nueva pregunta y se guarda como tal en la base de datos relacional.

# Forma manual de capturar respuestas

En Discord hay varias formas en las que un usuario puede responder a una pregunta:

1. Responder con una "respuesta citada"
2. Responder con una mención
3. Responder sin mención ni referencia
4. Responder en hilos

## Responder con una "respuesta citada"

En este caso, Discord crea una referencia explícita al mensaje original al que se está respondiendo (almacena el message\_reference que apunta al message\_id de la pregunta original).

Entonces, en la base de datos relacional se tendrían preguntas con message\_id, de manera que podría asociarse la respuesta con la pregunta mediante ese campo. Luego, se guarda la respuesta con su id\_pregunta correcto.

## Respuesta con Mención (sin citar el mensaje original)

Si alguien menciona a un usuario, pero no usa la función de "respuesta citada", no se tiene el message\_reference.

El mensaje tiene el o los user\_id

* Si hay una sola mención se podría buscar en la base de datos la pregunta más reciente realizada por esa persona. Si hay más de una pregunta reciente, se podría usar búsqueda semántica en ChromaDB para encontrar la más probable.
* Si fueron muchos los usuarios mencionados, buscar entre ellos cual fue la pregunta más reciente o hacer búsqueda semántica para encontrar cuál pregunta está más relacionada con esa mención.

Por lo tanto, en la base de datos relacional debería tener el autor y la fecha en que se hizo la pregunta.

## Responder sin mención ni referencia

Las respuestas sin mención ni referencia suelen ser respuestas independientes, que no hacen referencia directa ni a la pregunta original ni a ningún usuario en particular.

En este caso, se debe hacer una búsqueda semántica para encontrar la pregunta más relacionada.

## Responder en hilos

En Discord, se puede crear un hilo a partir de un mensaje. Todo lo que se escriba en el hilo pertenece a ese mensaje inicial.

Discord tiene un campo llamado thread\_id que indica que el mensaje forma parte de un hilo específico. Si el mensaje tiene un thread\_id, se puede reconocer la pregunta original que inició un hilo.

Todos los mensajes dentro de ese hilo deberían estar asociados con la misma id\_pregunta en la base de datos, ya que pertenecen a la misma conversación.

# Análisis de las técnicas manuales

Estas técnicas manuales para identificar preguntas y respuestas podrían tener muchos falsos positivos o negativos.

Además, en este chatbot se está buscando precisión en la identificación de pregunta y respuestas para proporcionar los mejores mensajes ante las consultas de los alumnos.

Por lo tanto, se ha elegido utilizar Mistral (un modelo grande de lenguaje) para analizar si un mensaje en Discord es una pregunta o una respuesta.