1.Introducción

1.1 Motivación y Objetivos

1.2 Marco de trabajo

1.3 Estructura de la memoria

2.Revisión de casos de uso

3.Análisis del problema

4.Preprocesamiento y análisis con Bert

5.Experimentos con Bert

6.Conclusión

**Utilidades Google Bert**

**Question Answering**

**Utilidades y Aplicaciones del Question Answering**

Uno de los principales enfoques, es la Atención al cliente.

Usado principalmente para la implementación de ChatBots, Asistentes Virtuales y/o Agentes de diálogo.

Todos estos modelos, buscan como objetivo, la obtención de respuestas claras y sobretodo coherentes.

Además de la Atención al cliente, este modelo, también sirve para la extracción de contenido clave en un texto.

**DataSet**

El modelo, ya ha sido previamente pre-entrenado con el SQuAD (Standfor Question Answering Dataset) dataset, existen dos versiones,

SQuADv1

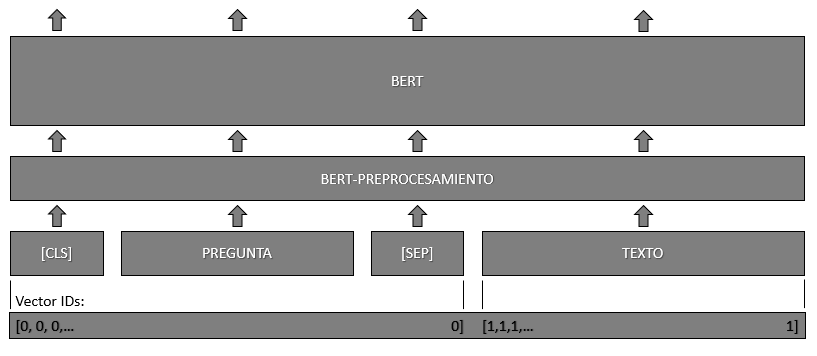
en la versión SQuADv1, podemos encontrar aproximadamente 100.000 preguntas, todas con respuesta.

SQuADv2

Mientras que en la versión SQuADv2, encontramos las 100.000 preguntas de la versión SQuADv2 a las que se suman aproximadamente 50.000 preguntas sin respuesta (reformuladas por los trabajadores para que puedan tener respuesta)

**Preprocesamiento**

Para analizar el texto y responder a las preguntas, el modelo, recibe tanto la pregunta como el texto, separados por un token especial que indica la separación.



**Interpretación de las Outputs**

El modelo, se ejecuta dos veces en paralelo, obteniendo dos vectores del tamaño equivalente a la longitud de la entrada, es decir,

Longitud de la salida = [3 + longitud(Pregunta) + longitud(Texto) ], donde el 3 se justifica por los tokens especiales que se añaden.

La primera Output, hace referencia a las puntuaciones para ser el token inicial, mientras que la segunta, para ser el token final.

Si obtenemos el máximo, obtendremos el índice con el token que mayor probabilidad ha obtenido.

Destacar, que el índice del token inicial, nunca ha de ser mayor al del token final.

La respuesta a la pregunta, será la comprendida en el rango: [Token\_Inicial, Token\_Final]

