**Methodology**

**Fine-tuned SBERT Clarification model**

A fine-tuned **SBERT (Sentence-BERT) model** is incorporated for ranking clarifying questions that handles the queries which cannot be resolved by BM25 on its own. The SBERT model that is used is a modification of the BERT(deep learning language model) that captures the semantic meaning of the sentences and relationships with other sentences. “Sentence-BERT (SBERT), a modification of the BERT network using Siamese and triplet networks that is able to derive semantically meaningful sentence embeddings.” **(Reimers & Gurevych, et al.,2019).** Sentence that now can be directly compared with reach other using the cosine similarity.A pretrained model is finetuned MiniLM-based SBERT Model called (Hugging face all-MiniLM-L6-V2) which is already trained for general task like sentence similarity. Which then trained to the qulac dataset. Each query and each clarification question got converted into vector of size 384 dimensions in this the similar text(query-clarification) comes closes whereas the dissimilar text vectors far apparat. By this space the relevant query-clarification have high cosine similarity, that enables a effective semantic matching of a query to its most suitable clarification.

**Training**

For training SBERT model was fine tuned to on a custom dataset of 196 ambiguous query -clarification enabling it to learn semantic relationships between the ambiguous users’ query and their most contextually appropriate clarifying question. A sentence transformer framework has been used with hugging face as transformers library in order to leverage the pretrained model and fine tune according to the task in hand. “Transformers is designed to mirror the standard NLP machine learning model pipeline: process data, apply a model, and make predictions”. [Transformers: State-of-the-Art Natural Language Processing](https://aclanthology.org/2020.emnlp-demos.6/) (Wolf et al., EMNLP 2020). **MultipleNegativesRankingLoss** was applied during the training because how it handles the training with only positive (pairs that are good match) without explicitly having a need for negative pair. This maximizes the cosine similarity of the true pair query-questions and while minimizing the similarity of other pairs. The model was fine tuned with 3 epochs and with a batch size of 16, which sufficient for the convergence of datasets (196 query-clarification pairs). The model’s architecture remains consistent with it dual- encoder nature. In which both the query -clarification pairs are encoded with mean pooling and L2 normalisation which is default in most SBERT model.

**Functional Pipeline:**

**1.Input** JSON data containing (Query-clarification) pairs.

**2. Training**- **The SBERT** model fine tuning first. These pairs were passed into SentenceTransformer model (pretrained all-miniLM-L6-v2) model after converting them to training examples. Using **MultipleNegativesRankingLoss** function. The model is trained for 3 epochs with a batch size of 16. The pairs with semantically similar query -clarification pairs were brought closer, which enables the use of cosine similarity to rank the best clarification question for the given query.

**3. Result**- The top Ranked clarification questions (with high similarity) where sent to RASA. Which then were presented to users as follow up question.

**Reference**

1. **Reimers, Nils and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.” *Conference on Empirical Methods in Natural Language Processing* (2019).**
2. [**https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2**](https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2)
3. **Sentence Transformers Library:**[**https://www.sbert.net/**](https://www.sbert.net/)
4. **[Transformers: State-of-the-Art Natural Language Processing] (https://aclanthology.org/2020.emnlp-demos.6/) (Wolf et al., EMNLP 2020)**