Asking Clarifying Questions for Conversational Search

Ansh Bisht and **Jordan Dickson** and **Jack Harrison** and **Ruben Lazell** and **Andrew Taison**
School of Computing, Engineering and the Built Environment,
Edinburgh Napier University
Matric Numbers: 40527530, 40545300, 40537035, 40679914, 40538519

Abstract

Short abstract.

1 Introduction

Brief context of problem - why clarifying Q's matter. What is the goal. Structure of the paper.

2 Related Work

Existing systems, prior research etc.

3 Methodology

3.1 System Overview

The system for the clarification model was constructed through a series of key components:User queries with a question for the system, RASA for application control of the entire system, Best Matching 25 (BM25) model for matching the queries on the Wikipedia Dataset, and SBERT for constructing Sentence Similarity Scores to rank the best results. The dependencies and relationships of these components are illustrated below.

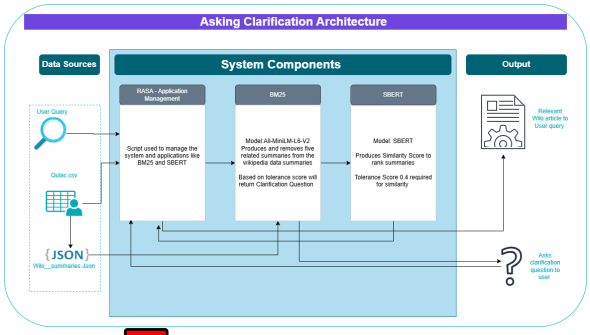

Figure 1: Enter Caption

4 Qulac Dataset

The Qulac data was one of the primary datasets for this system. One of the initial steps was to appropriately shape the data to work with the other elements within our system. The transformations for cleaning this dataset were performed initially in Python using the Pandas library, and the resulting CSV was stored within a Google Drive directory.

4.1 Original Qulac Dataset

A dictionary with query keys and a list of relevant facet descriptions (user intents).

4.2 Clean Qulac Schema

Table 1: Dataset Column Descriptions

Column Name	Type	Description
Index	String	Facet index based on word string.
Category Name	String	Name of the broader category of the
Facet Description	String	Detailed description of the facet.
Col 1	String	Response 1.
Col 2	String	Response 2.
Col 3	String	Response 3.
Col 4	String	Response 4.
New Index	Integer	Row index counter.

One of the most important features of our Qulac Dataset was the facet descriptions, as they were used to derive the Wikipedia Summaries.

5 Wikipedia Summaries

The Wikipedia summaries were generated using a series of GET responses contained within a Python script on the Wikipedia API. The aim was to generate five summaries per index – topic of the facet – relating to the user's query.

Once completed, the Wikipedia Summaries were shaped into the cleaned dataset above. These summaries were required for BM25 later in the project. In a larger project, the Wikipedia API could be used to return the entire articles, but this would come at a cost of time, compute, and complexity.

Figure 2: Wikipedia Summaries Generation

5.1 RASA

RASA is the machine learning framework used for building this system. It is used to manage the dialogue by deploying the suitable models. For this project, RASA deploys BM25 and SBERT, allowing the system to function cohesively.

5.2 Best Matching 25 (BM25)

BM25 was the key AI model in matching user queries to the relevant Wikipedia summaries within the system. The model used was from Huggingface and constructed via a Python script.

The documents used were the Wikipedia Summaries from the generated CSV. The result was that the system determined the need for clarification based on a tolerance score of 0.40 for similarity. A ranking was also performed using all-MiniLM-L6-v2 from HuggingFace to select the best matching summary.

5.3 SBERT

The final step was determining the suitability of the top-ranked result. A specialized version of BERT called SBERT from Huggingface was utilized.

Our SBERT model was fine-tuned on `sbert_training_data.csv`, which contained two features: context and question. Each context is a formatted string combining the initial query (tagged with `[Q]`) and the next element (a question and its answer tagged with `[A]`).

The model applied a similarity score with a tolerance threshold of 0.4. If the context and question's similarity score was above 0.4, the result was deemed suitable; otherwise, it was not ranked. (Aliannejadi et al., 2019)

7 Evaluation

How did we check the system works. Describe testing process.

8 Results and Discussion

Sample outputs. Summary of system behaviour. What worked well, what didn't? Possible improvements.

9 Conclusion

start here.

Limitations

Required by ACL format, and should be AFTER conclusion. Discuss honest limitations of the work.

Ethics Statement

Required by ACL format. Could just be a sentence or two. Explicit ethics statement on the broader impact of the work, or other ethical considerations.

References

Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 475–484. ACM.

A Appendix

Possibly not needed.