Textual Question Answering Systems- Frequently Asked Questions Retrieval

Submitted in partial fulfillment of the requirements of CS F366

By

Miloni Mittal 2017A3PS0243P

Aman Narsaria 2018B3A70743P

Under the supervision of:

Dr. Yashvardhan Sharma

&

Ms. Vijay Kumari

Birla Institute of Technology and Science Pilani, Pilani Campus

May 2021

Contents

Chapter 1

Introduction

1.1 Background

Building a Question Answering System is an important problem statement in the field of Natural Language Processing. It involves extracting the most relevant information from an abundance of information which may be classified into relevant, useful or irrelevant. Due to the information overload with respect to quantity and categories, searching for the relevant answer to the posed query is of utmost importance.
A question answering system can be built for two domains-

1. **Open Domain** There is no boundary on the category of content that is referred to for extracting the answer. *Example: Google search engine, Yahoo search engine*

2. **Closed Domain** There is a certain boundary on what queries this type of system can answer. This type of system restricts to a particular category of questions and in some cases even answers by referring to a document. *Example: QA services on various business websites, Solution providers for school textbooks*

Speed of deriving the answers to the user query plays a major role in such systems. This demands for a need to find methods that can extract relevant answers to the queries in the fastest manner possible. One method to do this is to look into the database of already asked and answered questions or FAQs and provide the top most relevant answers from that list.

1.2 Motivation

A question answering system is capable of working with natural language. It provides the user with the most relevant information in the least possible time. The challenges of working with such systems, usefulness of question answering system and having hands-on experience motivated us to take up this project.

Many questions have usually been answered in the FAQ section of a QA system. But the vast-ness of such a database might make it difficult for a user to search for the particular relevant query. Thus, answers to a user query posed to a closed domain QAS can be derived using the existing FAQ database. By checking if a similar question exists in the database, a faster and more efficient QAS can be built which does not need to access the original database for every query. We want to focus on the highlighted part of Figure 1.1

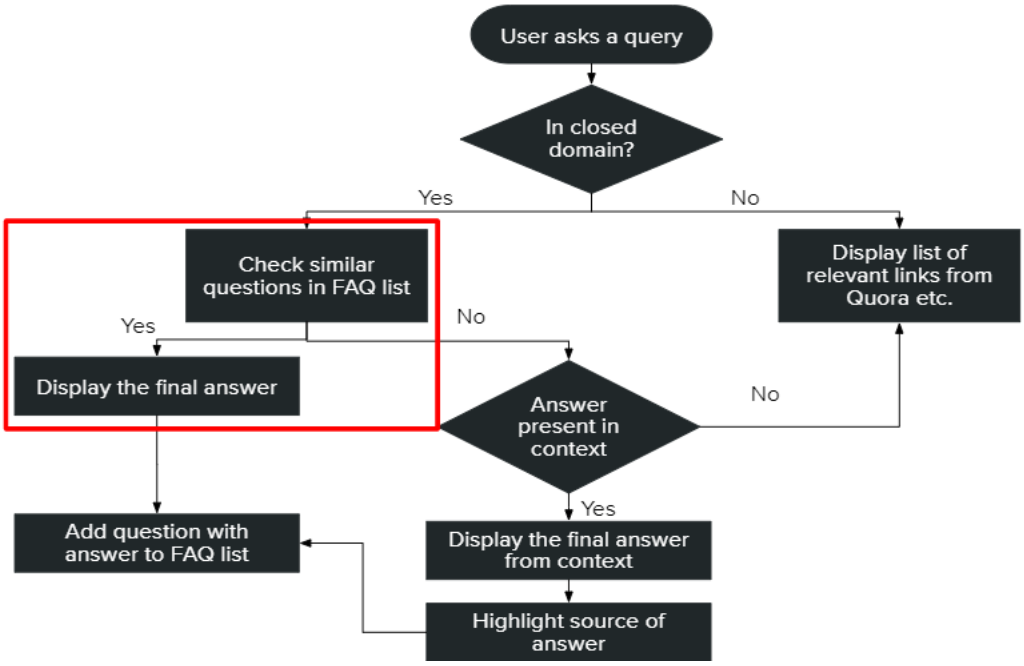

FIGURE 1.1: Flowchart for a Question-Answering system. We are focusing on the red highlighted part.

1.3 Objective

The objective of this project is to focus on the FAQ retrieval of the Question Answering System. We want to retrieve questions from the FAQ database which are relevant to the user query, thus providing answers to the user. Initially, this problem statement was solved only by referring to the similarity between the FAQ question (referred to as Q henceforth) and user query (referred to as q henceforth). But in recent times, researchers are leaning towards a more robust modelling involving FAQ answers (referred to as A henceforth) as well in similarity comparison (as shown in Figure 1.2). This method tends to give better results as lexical gaps in either comparisons (q-Q or q-A) can be compensated by the other.

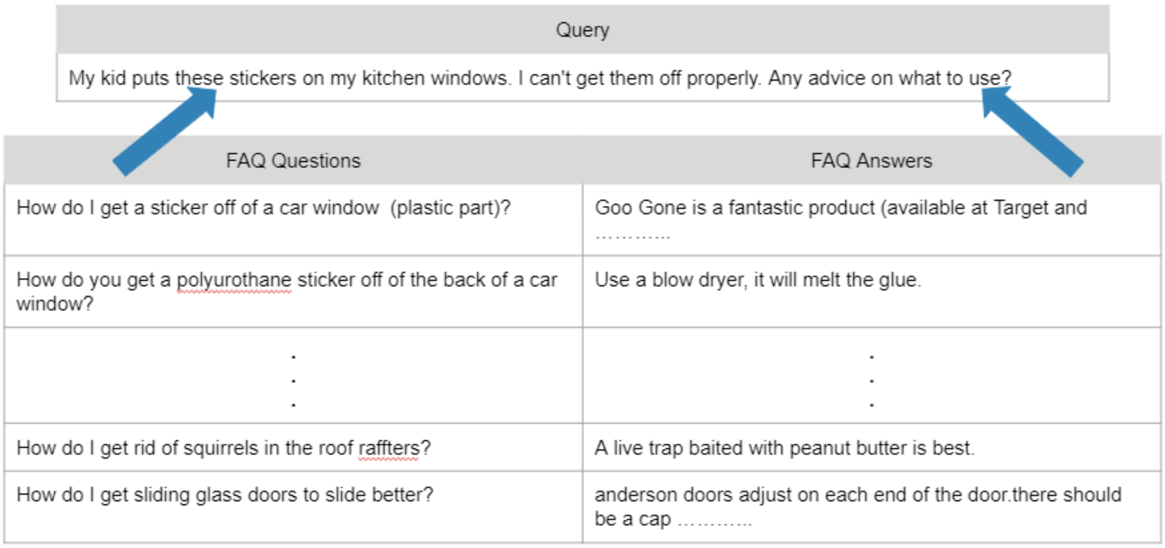

FIGURE 1.2: A diagrammatic representation of how queries will be compared with the FAQ question and FAQ answers

Chapter 2

Related Work

This section gives a concise review of the approaches reviewed.

2.1 Language Models

2.1.1 RNN's

The use of RNNs for QA has been explained based on their commendable performance on the most-basic datasets. They have hidden layers which maintain relationships with the previous values, thus giving them the ability to model long span dependencies. However, more complex datasets require significant feature engineering and hyperparameter tuning to achieve decent results [8].

2.1.2 BERT

Bidirectional Encoder Representations from Transformers [1] is one of the recent NLP models which has accomplished state-of-the-art results in a wide variety of NLP tasks, including Question Answering (SQuAD v1.1), Natural Language Inference (MNLI), and many others. BERT makes use of Transformer, an attention mechanism (as shown in Figure 2.1) that learns contextual relations between words (or sub-words) in a text.

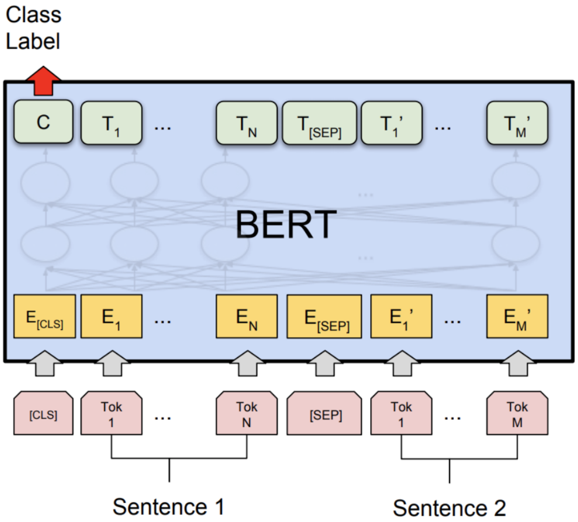

FIGURE 2.1: The BERT model

2.1.3 SBERT

Sentence-BERT [6], is a modification of the pretrained BERT network that uses siamese and triplet network structures (as shown in Figure 2.2) to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity. This reduces the effort for finding the most similar pair from 65 hours with BERT / RoBERTa to about 5 seconds with SBERT, while maintaining the accuracy from BERT.

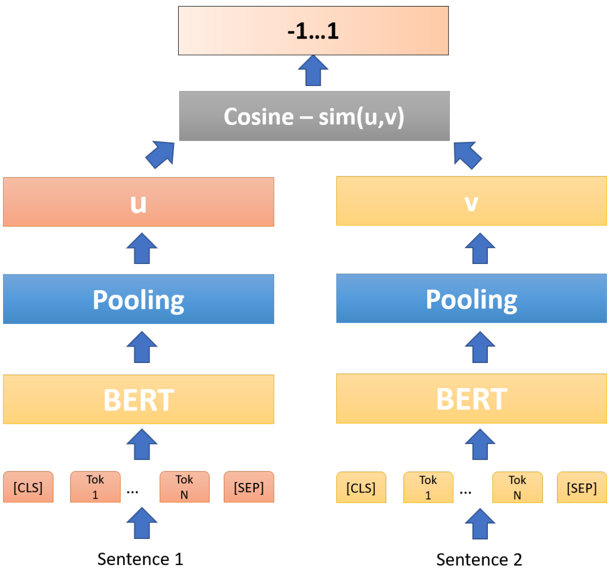

FIGURE 2.2: The SBERT model

2.1.4 DistilBERT

DistilBERT is a fast, cheap and light transformer based model based on the BERT architecture. This model is obtained using knowledge distillation during the pre-training phase to decrease the size of the BERT model by 40%. This model has 40 % less parameters than the original bert model but it preserves over 95 % of BERT's performance as mentioned in the paper [10].

2.1.5 XLNET

It depends on a summed up autoregressive pertaining strategy that empowers learning bidirectional settings by maximizing the normal probability over all permutations of the factorization order and consequently beats the restrictions of BERT [5].

2.1.6 SGNET

Syntax-Guided Machine Reading Comprehension [13] takes into focus the effective linguistic modeling of lengthy passages to get rid of the noises. The SG-Net model makes use of a context-aggregation mechanism for better representation of the linguistic dependence in the input sequence. This helps in speeding up of the model by significant amounts and creating a focus on the syntactical importance of specific words.

2.2 FAQ Retrieval Models

To retrieve the relevant FAQ pairs, many techniques have been explored but the most recent one includes q-Q and q-A similarity. The major hurdle in this technique is the unavailability of a training dataset for q-Q. Whereas, q-A training dataset can be easily built from the FAQ pairs. To tackle this problem, a wide variety of techniques are applied to either build the dataset or use other unsupervised techniques.

2.2.1 TSUBAKI for q-Q similarity and BERT for q-A relevance

[9] proposes a supervised technique for FAQ retrieval. It leverages the TSUBAKI model [11] for retrieving the q-Q similarity score. It is an unsupervised information retrieval system which is based on the OKAPI BM25 model [7]. For obtaining relevant q-A pairs, the BERT model is used.

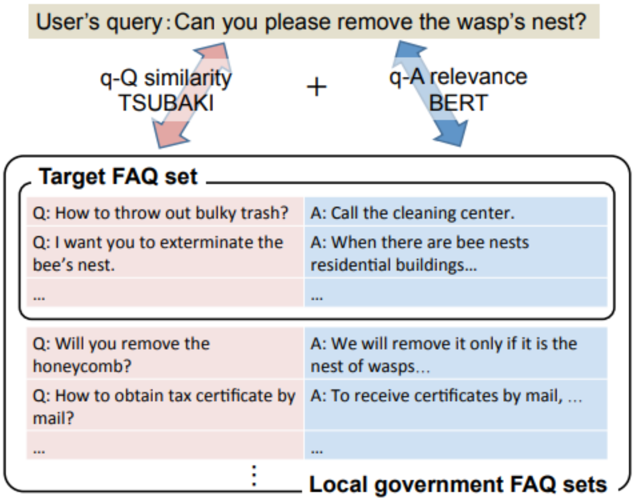

FIGURE 2.3: Unsupervised model for FAQ retrieval using TSUBAKI and BERT [9]

2.2.2 BERT model for both q-Q and q-A similarity

This model [4] uses BERT for training q-Q and q-A models. The lack of q-Q training dataset is compensated by a novel technique which generates question paraphrases. For the re-ranking, it uses elastic search, passage re-ranking and finally ranks on the bases of q-A and q-Q similarity.

2.2.3 FAQ Retrieval Using Attentive Matching

[2] proposes an attention mechanism model for FAQ Retrieval. It compares various aggregation methods to effectively represent query, question and answer information. It is observed that attention mechanisms are consistently the most effective way to aggregate the inputs for ranking. The use of Attentive matching in FAQ retrieval eliminates the need for this feature engineering and effectively combines both query-question and query-answer representations.

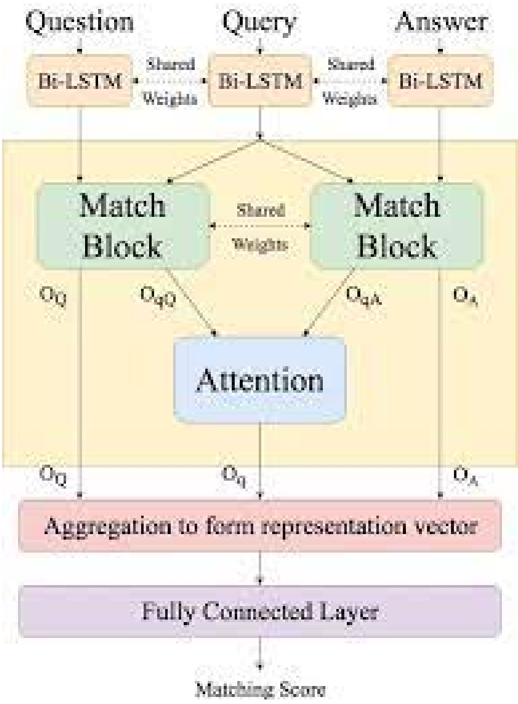

FIGURE 2.4: FAQ retrieval using Attentive Matching [2]

Chapter 3

Proposed Techniques and Algorithms

3.1 Preprocessing the data

The pre-processing steps for our models consists of:

1. **Lowercase**: The first step was to make the FAQ pairs and queries into lower case.
2. **Removing punctuations**: All punctuation marks were removed.
3. **Removing stopwords** : All stopwords were removed.
4. **Removing numbers** : All question numbers were also removed.

The SBERT and DistilBERT models used pre-processing steps 1, 2 and 4. The BM25 models used pre-processing steps 1, 2, 3 and 4. The BERT models used pre-processing steps 1 and 2.

3.1.1 Building the training dataset for q-A model

The original dataset consists of pairs of questions and answers from the FAQ database and queries with a list of questions from FAQ database that match the query. The final q-A model should be able to give a similarity score of whether the given answer matches the query or not. To build it we need to fine tune SBERT with a dataset that contains:

- (Q,A) the matching FAQ pair
- (Q,A') an FAQ question with a non-matching answer

This was done by randomly selecting A' for every question. The (Q,A') pairs were labeled as 0 whereas the (Q,A) pairs were labeled as 1 as shown in Figure 3.1.

FaqQuestion	Answer	Label
how do you seal leaking dorme	fixing a leaky roof especially somethi	1
how do i connect a us water filt	lots of elbow grease wash down the	0
how do i install front door spea	carefully remove all the screws from t	1
how do i change rear brakes or	couldnt find exactly your vehicle but	0
how do i grease my slip yoke g	i own a repair shop and the clunk yo	1
how do you replace a 3 way sw	go to your local hardware store hom	1
how do i fix a broiler in my gas	water pumps last less than 100k mile	0
how do you get to the actual sp	check with the owners manual also	1
how to wire hbl insulgrip twist l	i m not sure what you are trying to wi	1

FIGURE 3.1: A sample of the dataset used for training the q-A model

3.1.2 Building the training dataset for q-Q model

The model built with this dataset should be able to give a similarity score of whether an FAQ question matches the query. The training dataset was derived from the model proposed in [4]. Here is the link to the dataset used. The label 1 was assigned to the corresponding question-paraphrase pairs. The label 0 was assigned to the question pairs that did not match. The second half of the dataset was build by random selection of a question from the FAQ database.

3.2 Models

3.2.1 SBERT

SBERT for query-answer (qA) comparison was trained in two ways -

1. Taking 1:1 ratios in A vs A' ratio for the dataset as described in Section 3.1.1.
2. Taking 1:5 ratios in A vs A' ratio for the dataset as described in Section 3.1.1.

Two variants of SBERT for query-question (qQ) comparison have been tried. SBERT encoding was directly used to obtain the similarity score for query and FAQ questions. Fine-tuned SBERT was built by training the SBERT model using the dataset described in Section 3.1.2.
For the query-answer (qA) comparison model, bert-base-uncased was used to fine tune the SBERT model. Dataset described in Section 4.1 was used.

3.2.2 DistilBERT

For query-answer (qA) model, distilbert-base-uncased was used to fine tune the model. For query-question (q-Q) comparison model also, distilbert-bert-uncased was used to fine tune the model and obtain sentence embeddings. The datasets described in Section 3.1.1 and 3.1.2 were used for training.

3.2.3 BM25 qQ

The corpus is built using the pre-processing methods as mentioned in Section 3.1.2 on the FAQ questions. Then the BM25 model is applied using the rank-bm25 library. The top 100 results are retrieved and the performance metrics are calculated for them.

3.2.4 BM25 q(Q+A)

Each corresponding FAQ Question and Answer is concatenated and is represented as Q+A. The corpus is built using the pre-processing methods as mentioned in Section 3.1 on the FAQ Q+A. Then the BM25 model is applied using the rank-bm25 library. The top 100 results are retrieved and the performance metrics are calculated for them.

3.2.5 BM25 q(Q+A) + BERT qA

The BERT model is trained on triplets (question, corresponding answer, non-corresponding answer) to understand the intricacies of matching and non-matching answers. The learning rate is 2e-5 and number of epochs is 3.

Example of triplets used in training BERT qA model

Question - How do you change an alternator?
Answer - Depending on what model car you have it will require different steps most libraries have manuals for these operations chilton's is probably the best if you can't find one at the library I'm sure you can buy one for your car online good luck
AnswerDash - You need to flush out the water heater with a garden hose it is probably filled with little rocks the inlet is probably at the bottom of the tank

Top 100 FAQ pairs are picked using BM25 Q+A. The encoding of the answers of these 100 FAQ pairs is found and compared with the query encoding using cosine similarity. The FAQ pairs are re-ranked based on these cosine similarity scores.
The relevance of the retrieved FAQ pairs is cross-checked with the relevance score in the dataset and the performance metrics are calculated accordingly.

3.2.6 BM25 q(Q+A) + BERT qQ

The BERT model is trained on triplets (question, paraphrase, non-matching question) to understand the intricacies of matching and non-matching questions. The learning rate is 2e-5 and number of epochs is 3.

Example of triplets used in training BERT qQ model

Question - How to get rid of garbage disposal odor?
Paraphrase - How do I clean the disposal and how do I get rid of the smell of paraffinic garbage?
QuestionDash - How do you fix a heater for on a van?

Top 100 FAQ pairs are picked using BM25 Q+A. The encoding of the questions of these 100 FAQ pairs is found and compared with the query encoding using cosine similarity. The FAQ pairs are re-ranked based on these cosine similarity scores.
The relevance of the retrieved FAQ pairs is cross-checked with the relevance score in the dataset and the performance metrics are calculated accordingly.

Chapter 4

Datasets Used

4.1 FAQIR Dataset

We use FAQIR [3] dataset for evaluation.The FAQIR dataset was derived from the "maintenance & repair" domain of the Yahoo! Answers community website. It consists of 4313 FAQ pairs and 1233 queries with corresponding manually annotated relevance judgements. The judgements are described as: 1- relevant, 2- useful, 3- useless and 4- irrelevant. Each query has at least one FAQ-pair annotated as "relevant". However, it is possible for a FAQ-pair to be irrelevant for all queries. The original dataset can be seen in Figure 4.1, 4.2 and 4.3.

4.2 Other Datasets

There are other datasets such as StackFAQ and COUGH Dataset [12] which provide FAQ Question and Answers along with the queries. StackFAQ holds some amount of ambiguity with respect to what is to be treated as FAQ pair and what is to be treated as query. If this ambiguity can be solved, this dataset can be utilised. The COUGH dataset is a multilingual dataset and can be explored too.

```
<Query>
  <Expansions />
  <Author>mladen</Author>
  <original>How do I get a sticker off of a car window  (plastic part)?  Nail polish remover seems risky?</original>
  <infneed>Inf. potreba -- kako skinuti naljepnicu sa staklene povrsine</infneed>
  <exWords>
    <string>sticker</string>
    <string>glass</string>
    <string>window</string>
    <string>mirror</string>
  </exWords>
  <QueryGroupID>0</QueryGroupID>
  <QueryID>0</QueryID>
  <QueryString>My kid puts these stickers on my kitchen windows. I can't get them off properly. Any advice on what to use?</QueryString>
  <relDocs />
  <Fold>0</Fold>
</Query>
```


FIGURE 4.1: Snapshot of the query dataset from the FAQIR dataset


```
<qaPair>
  <Id>5173</Id>
  <Question>how to fix scratched mirror? </Question>
  <Answer>.if the scratch isn't too big it can be polished with toothpaste. Just apply a small
  <Categories>
    <string>maintenance &amp; repairs</string>
    <string>home &amp; garden</string>
    <string>maintenance &amp; repairs</string>
  </Categories>
  <IncludeAnswerInGetWords>true</IncludeAnswerInGetWords>
  <IncludeQuestionInGetWords>true</IncludeQuestionInGetWords>
</qaPair>
```


FIGURE 4.2: Snapshot of the FAQ pairs dataset from the FAQIR dataset


```
<IRCandidate>
  <Id>10006</Id>
  <Annotations>
    <Annotation>
      <Val>1</Val>
      <Annotator>djurozlikovski</Annotator>
      <AnnotationTime>2015-10-10T17:22:57.7533266+02:00</AnnotationTime>
    </Annotation>
  </Annotations>
</IRCandidate>
```


FIGURE 4.3: Snapshot of the relevance of FAQ to query mapping dataset from the FAQIR dataset

4.3 Reason for not using other datasets

There is a variety of other datasets available for the FAQ retrieval such as the Quora Question pairs. The reason for not utilising this dataset is that they don't contain data on the answer aspect of the FAQ; they only contain question pairs. The aim of the project is to also utilise the answers to retrieve relevant documents. Other datasets like SQuAD don't have descriptive answers which are necessary for q-A model training.

Chapter 5

Experiments and Results

5.1 Evaluation and Results

Various models and ranking techniques were explored. The following performance metrics have been used for the retrieval:

1. **Mean Precision at 5 (P@5)** is the measure of number of relevant documents in the top 5 retrieved documents. It helps to determine how many relevant document are ranked in top 5. The more documents in top 5, the better the IR system is.

2. **Mean Average Precision (MAP)** is a measure of whether all of the relevant documents get ranked highly or not. It is needed because a relevant document being retrieved but present lower in the list would not be very useful for a user entering his/her query.

3. **Mean Reciprocal Rank (MRR)** is a measure of the rank at which the first relevant document occurs in the retrieved documents.

The metrics obtained for different models has been shown in Table 5.1. We see that best results (P@5 = 0.42, MAP = 0.51, MRR = 0.69) have been obtained using the BM25 q(Q+A) + BERT qQ model. An example of retrieval of FAQ questions is shown in Figure 5.1 and 5.2.

TABLE 5.1: Performance metrics

Model	Ranking Method	P@5	MAP	MRR
1:1 qA training SBERT	(top qA, sort qQ)	0.14	0.32	0.33
1:5 qA training SBERT	(top qA, sort qQ)	0.19	0.35	0.37
1:5 qA + qQ training DistilBERT	(0.2*qA score+0.8*qQ score)	0.18	0.30	0.40
BM25 qQ training	(top 100 FAQ Q)	0.30	0.38	0.57
BM25 q(Q+A) training	(top 100 FAQ Q+A)	0.34	0.39	0.60
BM25 q(Q+A) + BERT qA training	(BM25 top 100 + rerank qQ)	0.27	0.32	0.52
BM25 q(Q+A) + BERT qQ training	(BM25 top 100 + rerank qQ)	0.42	0.51	0.69

```
User Query : My floor is awful. It's made of wood and it makes a lot of noise. Enough to wake someone up. Can we repair it somehow?
Retrieved Answer 1 : If you have a basement, or a crawl-space, you can drill screws from underneath (being careful not to use too long of a screw
Retrieved Answer 2 : the best way to do this is if you have a basement and you can access the floor joists.  if you can have someone walk around
Retrieved Answer 3 : Some people think that creaking hardwood floors are quaint.  If it is not loose, try sprinkling baby powder on the area, vac
Retrieved Answer 4 : Squeaky floors are fairly common, and there aren't many homes that do not have at least one. Although they're aggravating, t
Retrieved Answer 5 : Depends on the size of the holes, but small ones (>1/2" aprox) can be filled with a good quality epoxy.  If you have a hidde
[1, 1, 1, 1, 5]
User Query : Recently I've been unable to fully close the rear left window on my car, and I'd like to replace it. Can I do it myself?
Retrieved Answer 1 : Before you replace Ranger rear window, consider replacing with a sliding rear window available at most auto part stores. It c
Retrieved Answer 2 : I assume you need to pop the clip off the back side of the handle? Try using a shop towel held in both hands, you will eventu
Retrieved Answer 3 : Take the door panel off then look to see what is broke and replace that part it should work like new then
Retrieved Answer 4 : I'm assuming you're not asking about the windshield.Instructions with pictures for the front windows can be found here:http://
Retrieved Answer 5 : I replaced mine by popping out the inside panel of the door, then taking the replacement glass and inserting it in it's needed
[1, 5, 1, 1, 1]
User Query : I have a sliding patio door that keeps opening by itself. How can I fix it?
Retrieved Answer 1 : You can get a cat door but what I did was get a little skinny bungee cord at the hardware store and attached it to the scree
Retrieved Answer 2 : If I'm reading your question right, there may be a pretty easy remedy. To correct the hitting near the top of the jamb, remo
Retrieved Answer 3 : No, you shouldn't have to open the wall. Pocket door hardware varies but they are all made to maintenance. Try removing the
Retrieved Answer 4 : Cast_Chris, I'm not sure which type of door you have.  If the screen you have came with your door then you might have the ty
Retrieved Answer 5 : Ok Tessiecom,  If the door you have is not a "pre-hung" door then you're in for a challenge.  Although a pre-hung door can b
[1, 4, 2, 3, 4]
```

FIGURE 5.1: Top 5 FAQs retrieved for some of the queries using SBERT Model

```
User Query : How do I get rid of wine stains on a carpet.
Retrieved Answer 1 : When red wine is spilled onto your carpet, white wine can be your true companion. White wine will neutralize red wine and will make it easier
Retrieved Answer 2 : Hi mailensp, this website says that red wine doesn't stand a chance against these cleaning tactics.  Hope this helps, good luck.
Retrieved Answer 3 : the same steam cleaner that you would use on your carpet in your house. you might need speacial attachments for the upholstery and stuff.
Retrieved Answer 4 : From my experience, it's virtually impossible to completely get rid of mold and mildew from anything absorbent.  Given the toxicity of mold, I
Retrieved Answer 5 : Contact a hardwood floor restoration expert. I think it can be done through a stripping,bleaching, sanding and refinishing process. I wouldn't
[1, 1, 1, 1, 4]
User Query : How do I install an electrical outlet?
Retrieved Answer 1 : In order for a grounded outlet to work safely, it should be used with 3-wire cable and be grounded to the ground wire through the service pan
Retrieved Answer 2 : Just cut the wire(Breaker off) Get a junction box and make up the three blacks 3 whites and three greens(or bare) wires.  Make sure you make
Retrieved Answer 3 : Turn off the power to the outlet, undo the outlet and replace with a switch, simple, takes about 5 minutes at the most.
Retrieved Answer 4 : Hardtop Models Disconnect the negative battery cable. Remove the air conditioner electrical connector by accessing through the glove box. Rem
Retrieved Answer 5 : Sounds like you're just overloading the circuit. Put your heat on a dedicated breaker by itself.
[1, 1, 1, 4, 5]
User Query : How to remove rust?
Retrieved Answer 1 : well you can go 2 ways. remove or reform. If you want to completely remove and replace all rusted areas, it will be costly. Or you can have
Retrieved Answer 2 : vinegar and lemon juice mixed with a little table salt will take the marks out. Also put a little bleach on the spots, let stand for a few m
Retrieved Answer 3 : Try the product kaboom. It used to be advertised on TV. You can get it at walmart. Walgreens used to sell it but cost more than at walmart.
Retrieved Answer 4 : how about dont paint it or remove the rust......spray the screen with Pam cooking spray of even better rub it with lard or butter....it will
Retrieved Answer 5 : Lemon juice and baking soda. =3
[1, 1, 1, 1, 1]
```

FIGURE 5.2: Top 5 FAQs retrieved for some of the queries using BM25 q(Q+A)

5.2 Error Analysis

We observe that the BM25 q(Q+A) + BERT qQ model gives the best results among all other models. This is because the BM25 model focuses on lexicons in the corpus and the BERT model focuses on the semantic meaning. Hence, they complement each other. BM25 q(Q+A) works better than BM25 qQ because concatenating the answer with the question provides more scope for matching lexicons. Words present in the query may be absent in the FAQ question but present in the FAQ answer.

The BM25 q(Q+A) + BERT qA model does not work well in comparison to BM25 q(Q+A)

+ BERT qQ model. This is because the semantics of a query and answer are usually very different. And hence, BERT qA does not perform that well. It is also observed that on adding the BERT qA model to the BM25 q(Q+A) model, the performance worsens.

5.3 Web Interface

The web interface was built using HTML, CSS, JavaScript and Flask. The model implemented in the frontend is the BM25 q(Q+A) + BERT qQ model described in Section 3.2.6. The frontend consists of a home page Figure 5.3 where the user enters the query and sees the results. The result has two aspects, an answer (Figure 5.4) and a "People also asked" section (Figure 5.5) which lists 5 pairs of FAQ which are similar to the query. Another sample query has been shown in Figure 5.6 and 5.7. Click here for the web interface demo.

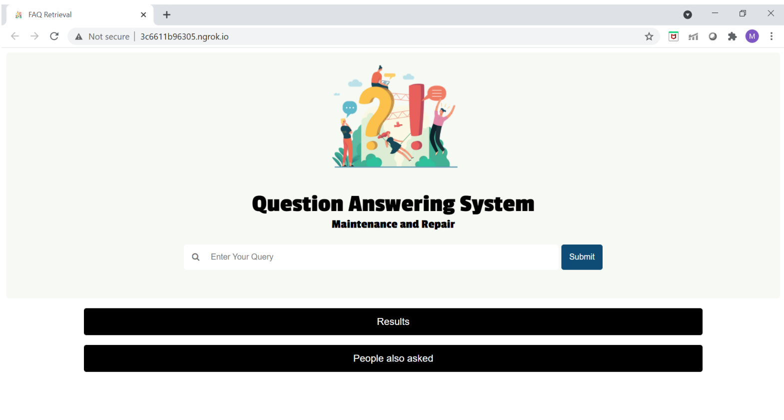

FIGURE 5.3: The Home Page

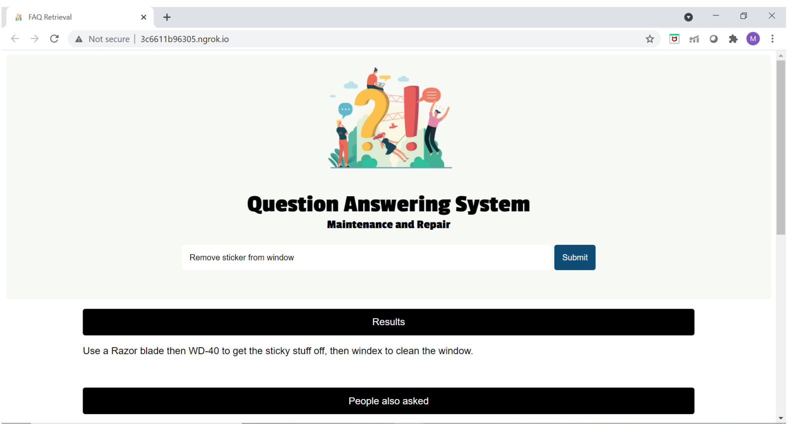

FIGURE 5.4: The Result

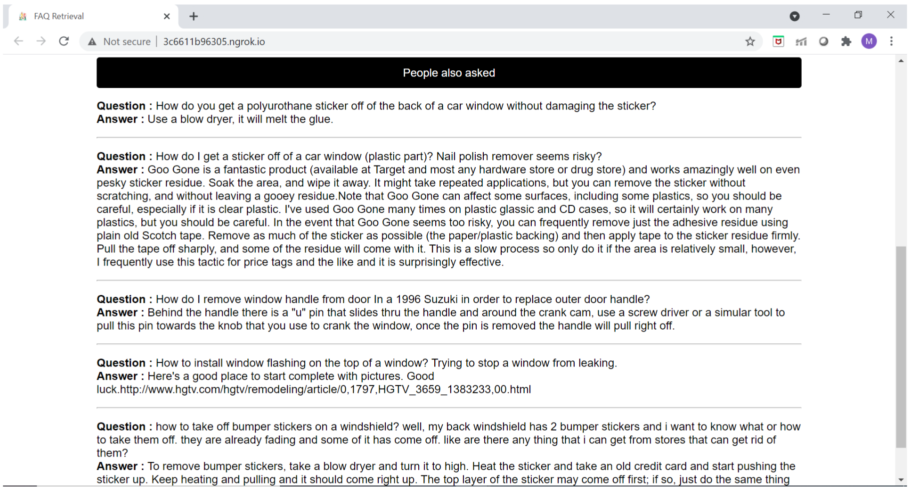

FIGURE 5.5: "People also asked" section

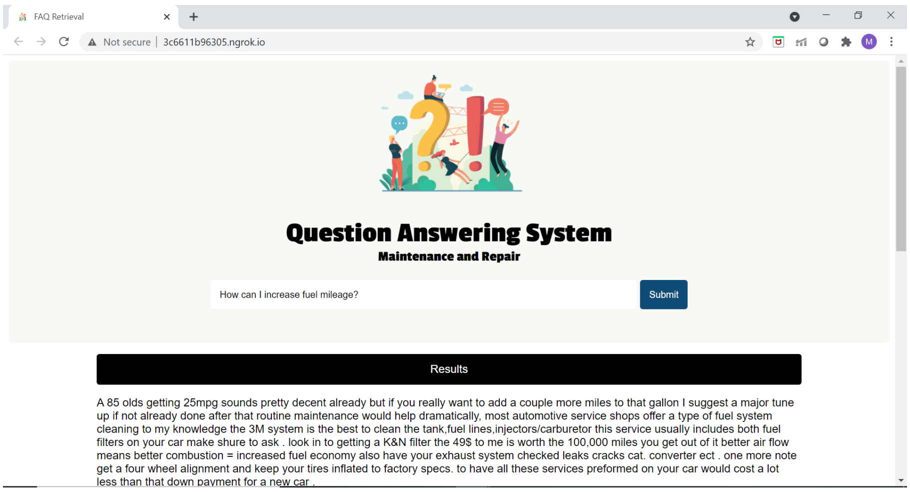

FIGURE 5.6: Answer for Query - "How can I increase fuel mileage?"

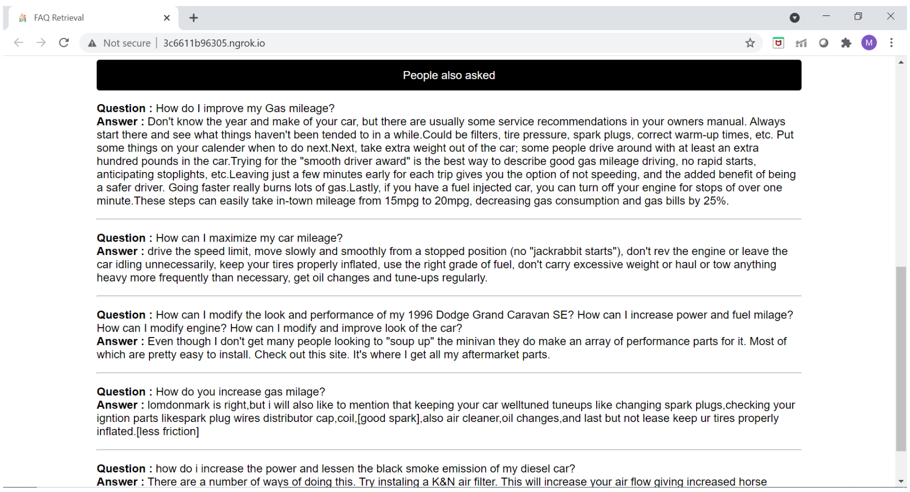

FIGURE 5.7: People also asked section for Query - "How can I increase fuel mileage?

Chapter 6

Conclusion and Future Work

We have successfully trained all our models on FAQIR dataset to evaluate the task of FAQ retrieval. We have also used various ranking methods as shown in Table 5.1.

We studied models based on BERT, TSUBAKI, BM25, Attentive Matching and taxonomy matching. We finally built a fusion model using the techniques from all models. Starting from vanilla SBERT model with a P@5, MAP and MRR of just 0.14, 0.32 and 0.33 respectively, we present our final BERT qQ + BM25 q(Q+A) model with P@5, MAP and MRR of 0.42, 0.51, 0.69 respectively. Our final model is a fusion of triplet trained BERT and BM25 ranking function as described in Section 3.2.6

We have also designed a web interface using HTML, CSS, JavaScript and Flask that uses our final FAQ model to fetch the most relevant answer and top 5 similar FAQ question and answer pairs.

Chapter 7

Work Update

7.1 Work completed

1. Learnt about the theoretical aspects of NLP and Q-A Retrieval Systems.
2. Carried out a literature review for FAQ models in Question Answering Systems in the field of information retrieval.
3. Learnt about the working and implementation of models like BERT, SBERT, BM25.
4. Experimented with various ranking techniques [weighted measures, re-ranking after initial retrieval] to rank top FAQ pairs.
5. Built a website using HTML, CSS, JavaScript and Flask and integrated our final model [BM25 q(Q+A) + BERT qQ training] with it.
6. Created an end-to-end website which gives top answer based on FAQ from the FAQIR dataset and 5 FAQ pairs that are similar to that category.

7.2 Future Work

1. Further improving the accuracy of our model by using alternative techniques.
2. Training and testing on other datasets like COUGH, StackFAQ.
3. Suggest better question framing (Like "Did you mean?" in Google).
4. Make a more generic FAQ system which caters to more than just one category.

Bibliography

[1] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[2] Sparsh Gupta and Vitor R. Carvalho. "FAQ Retrieval Using Attentive Matching". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR'19. Paris, France: Association for Computing Machinery, 2019, 929–932. ISBN: 9781450361729. DOI: 10.1145/3331184.3331294. URL: https://doi.org/10.1145/3331184.3331294.

[3] Mladen Karan and Jan Šnajder. "FAQIR – a Frequently Asked Questions Retrieval Test Collection". In: *Proceedings of the 10th edition of the Language Resources and Evaluation Conference, LREC 2016*. ELRA. 2015.

[4] Yosi Mass et al. "Unsupervised FAQ Retrieval with Question Generation and BERT". In: Jan. 2020, pp. 807–812. DOI: 10.18653/v1/2020.acl-main.74.

[5] B. Myagmar, J. Li, and S. Kimura. "Cross-Domain Sentiment Classification With Bidirectional Contextualized Transformer Language Models". In: vol. 7. 2019, pp. 163219–163230. DOI: 10.1109/ACCESS.2019.2952360.

[6] Nils Reimers and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: 2019. arXiv: 1908.10084 [cs.CL].

[7] Stephen E. Robertson, Steve Walker, and Micheline Beaulieu. "Experimentation as a way of life: Okapi at TREC". In: *Information processing & management* 36.1 (2000), pp. 95–108.

[8] G. Rohit, Ekta Dharamshi, and Natarajan Subramanyam. "Approaches to Question Answering Using LSTM and Memory Networks: SocProS 2017, Volume 1". In: 2019, pp. 199–209. ISBN: 978-981-13-1591-6. DOI: 10.1007/978-981-13-1592-3_15.

[9] Wataru Sakata et al. "FAQ retrieval using query-question similarity and BERT-based query-answer relevance". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019, pp. 1113–1116.

[10] Victor Sanh et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *CoRR* abs/1910.01108 (2019). arXiv: 1910.01108. URL: http://arxiv.org/abs/1910.01108.

[11] Keiji Shinzato et al. "Tsubaki: An open search engine infrastructure for developing information access methodology". In: *Journal of information processing* 20.1 (2012), pp. 216–227.

[12] Xinliang (Frederick) Zhang et al. "COUGH: A Challenge Dataset and Models for COVID-19 FAQ Retrieval". In: *arXiv preprint arXiv:2010.12800* (2020).

[13] Zhuosheng Zhang et al. "SG-Net: Syntax-Guided Machine Reading Comprehension". In: *CoRR* abs/1908.05147 (2019). arXiv: 1908.05147. URL: http://arxiv.org/abs/1908.05147.