Initial Lessons from Building an IVR-based Automated Question-Answering System

Pranav Bhagat, Sachin Kumar Prajapati, Aaditeshwar Seth, IIT Delhi

Introduction

FAQ retrieval, a form of QA systems where questions are matched against a database of FAQs to find an approximate answer

Kahi Ankahi Baatein (KAB):

- An IVR system on Sexual and Reproductive Health & Rights(SRHR)
- Manually operated QA programme(in Hindi) callers ask questions about SRHR

FAQ-based automated Question-Answering can be made to work on IVR systems, to service less-literate populations who prefer to use voice as an interaction modality

Problem Statement

We have a set **F** of FAQs with n question-answer pairs :-

$$\mathbf{F} = \{ (Q_1, A_1), (Q_2, A_2), \ldots\ldots (Q_N, A_N) \}$$

Note: Multiple questions can have the same answer

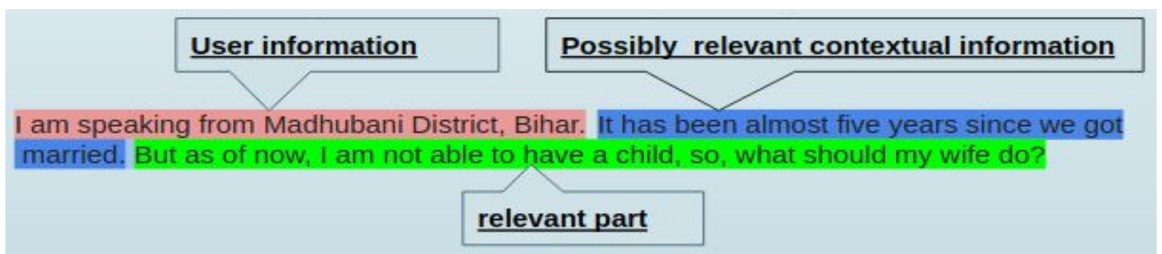
Now, given a user query **q**.

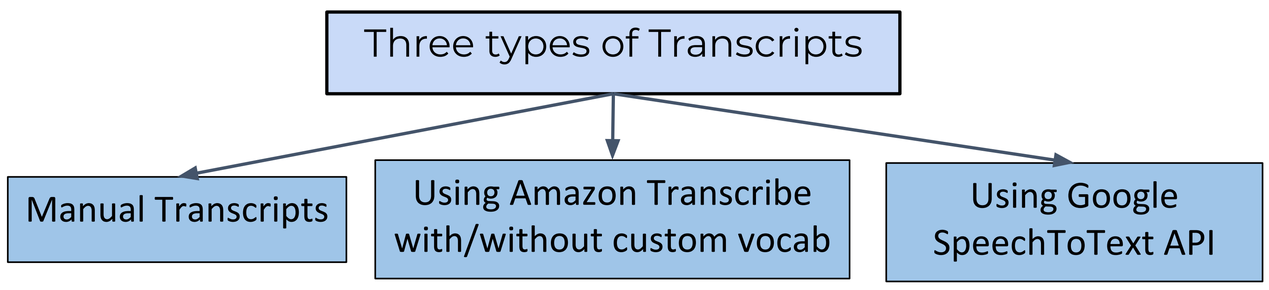
Task: To find an ordered subset **A** from the set F, such that pairs (Q_i, A_i) in A answer user query **q**.

Subset **A** should also be sequenced in order of their relevance to the user query **q**

KAB Dataset

- A collection of audio files of 90 answers and 516 questions
- A typical Question:

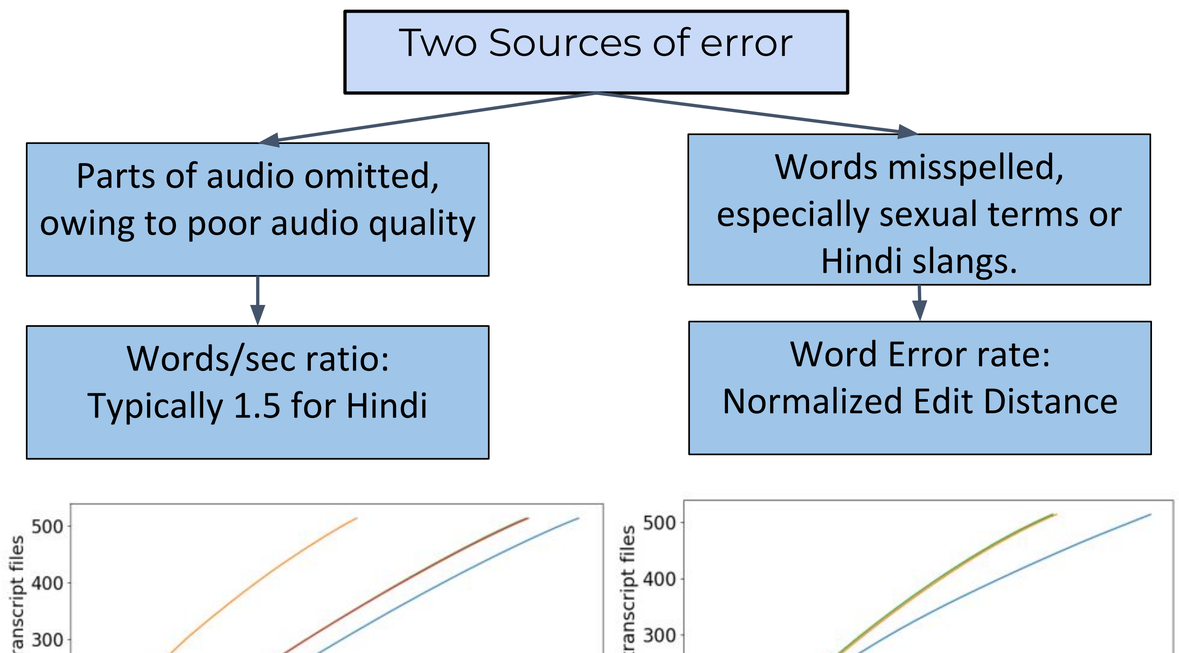




- Some of the incorrectly transcribed words* in Hindi.

सेक्स(sex) -> टेस्ट(Test) , सेट(Set)
यौन सम्बन्ध(Yaun Sambandh) -> जो समन(Joe Summon)
माहवारी(Mahavari)-> वहाँ भारी (vaham Bhari)
एचआईवी एड्स(HIV AIDS) -> एक भी एड(Ek Bhi ed)

*Words in red were transcribed correctly after listing them in a custom dictionary



(a) CDF for words per sec ratio of questions **(b) CDF for word error rate of questions**

Methodology

- FAQs organized as a set of (Q, A) pairs.
- New user query(q) is matched against this set to find appropriate answer.
- Matching is based on finding sentence similarity.
- Two methods used for sentence similarity:-
  - **Jaccard Similarity:** based on Jaccard index
  - **BERT_multilingual:** pretrained BERT model by Google

Results

Results produced for five different cases depending on the type transcriptions

	Models Used	Success Rate in top-5 results (SR@5): Exact Match	Approx. Match
Training and test data automatically transcribed	BERT	**0.143**	**0.121**
	Jaccard Sim.	**0.229**	**0.26**
Training and test data manually transcribed	BERT	**0.359**	**0.593**
	Jaccard Sim.	**0.394**	**0.463**
Only relevant portions of the questions are considered	BERT	**0.57**	**0.697**
	Jaccard Sim.	**0.557**	**0.629**
Training and test data is segmented into broad themes, search within a theme	BERT	**0.71**	**0.751**
	Jaccard Sim.	**0.733**	**0.787**
Training data transcribed manually, query transcribed using Amazon Transcribe	BERT	**0.593**	**0.643**
	Jaccard Sim.	**0.543**	**0.633**

Conclusions

- Not able to perform well, indicating the need for manually transcribed data
- Uses manual transcripts, works better, but unrealistic since query data cannot be manually transcribed
- Even better as it eliminates superfluous information from both the questions and answers
- Results further improve when the search space is restricted, implies an upfront IVR-based choice for the broad theme
- Performance in real-world scenario, as user query cannot be manually transcribed on the go