An Alternate Approach for Question Answering system in Bengali Language using Classification Techniques

Arijit Das

Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, Kolkata, West Bengal, 700032 INDIA
email- arijit.das@ieee.org

Abstract - Question Answering (QA) system is becoming more popular with the introduction of Virtual Agents and Chatbots. Medium of QA system is generally either text or audio. There are differences between search engine and QA system. Generally searching is based on keyword matching. In case of web search, list of URLs is ranked based on location, user history, search preference etc. Sophisticated algorithms like page-rank is also involved there. On the other hand, QA system does not work on keyword matching primarily. It's often possible that the query and the best answer have no term or a very small number of terms in common. QA system in English and other popular languages resolves the issues with the help of ontology, WordNet, machine readable dictionary etc. QA system in low resource languages suffers from lack of annotation, absence of WordNet, immature ontology. In this work, QA system in Bengali is developed using supervised learning algorithms. A collection of Bengali literatures, which was developed during TDIL (Technology Development of Indian Languages) project funded by Govt. of India, is used as the repository. Well known classification techniques like ANN, SVM, Naïve Bayes and Decision Tree are employed in this work. The system has achieved 84.33% accuracy to return the exact answer. It has achieved 97.13% accuracy to return the string containing correct answer. Unavailability of structured dataset and poor resources were the main challenges for this work. QA system in Indian languages especially Bengali is very much useful not only for chatbots or virtual agents but also for the e Governance and mobile governance in West Bengal and Bangladesh. QA system in mother tongue gives opportunity to more number citizens to interact with the administration. Though the system is designed aiming towards Bengali language but it can be tuned to work for any language with minimum modification.

Keywords*:* QA System; SVM; Naive Bayes; Neural Network; Decision Tree

(Received March 31, 2020/ Accepted June 9, 2020)

1. Introduction

Alexa, Siri, Google Assistant, Cortana have become the most common names in the digital world now a days. They are the virtual assistants (VAs) available in digital devices like desktops, laptops, mobiles, watches etc. These VAs interact with human beings in natural languages. The comunication may be through voice or text. The VA being asked gives reply based on knowledge and information available. Generally, they are connected with the internet and use the web repository as their

extended knowledge source. To give the correct response understanding the question is the primary step. Question Answering is seen as special kind of searching, where semantic knowledge is required to understand the system. Questions are not necessarily be English rather asking questions to the VAs in local languages are increasing. So, the virtual assistants are one of the latest agents to understand the trends of human computer interaction.

Question Answering System was first implemented in the BASEBALL and LUNAR project. BASEBALL was to answer the questions related to the popular baseball tournament. LUNAR was to answer the questions about the geological data of the Apollo Moon Mission. All the early QA systems were very much static. They gave fixed set of answers against the fixed set of questions. Research on QA system got the velocity from 2001 with the increased demand of automation of the customer care from the industry. All the QA systems were targeted for English primarily. Later research started for other popular languages.

The governments in various countries started to feel the increasing importance of QA system during implementation of various e governance and m governance projects. Major countries in the world are multilingual. So, necessity for the QA systems in national and regional languages increased. In India, Amazon has started to rollout R&D project to train Alexa in Hindi. Researchers have started to address the challenges of building QA system in various Indian languages like Bengali, Odiya, Punjabi, Tamil, Telegu, Malayalam, Kannar, Manipuri etc.

In this work, QA system in Bengali using classification technique is developed. A collection of Bengali literature, which was developed in the Technology Development of Indian Languages (TDIL) project funded by Ministry of Electronics and IT, Govt. of India, is used as the repository of the QA system. Details of the repository may be found in the TDIL website. There are eighty-six categories of text in the repository.

First the system is trained with the training dataset and next the system is tested with the test dataset or questions. Some examples of the questions are given below-

কোন সালে হ্যালির ধূমকেতু পৃথিবীর আকাশে দেখা দিয়েছিল ? (‘Kon sale Haleyr dhumketu prithibir akash e dekha giechhilo?’ or ‘In Which year Halley's Comet was shown in the sky of the Earth?’)

মিশরে কলেরা রোগ এর প্রাদুর্ভাব প্রথম কবে হয় ? (‘Mishor e kolera rog er pradurvab prathom kobe hoy?’ or ‘When there was the first outbreak of cholera in Egypt?’)

১৯০৮ সালে কে রসায়নবিদ্যায় নোবেল পুরষ্কার লাভ করেন ? (‘1908 sale ke rosayanbidyay nobel puroskar lav koren?’ or ‘Who got the Nobel Prize in chemistry in 1908? ’)

সাগর থেকে সরকারী হিসাবে রুদ্রনাথের দরত্ব কত ?(‘sagar theke sarkari hisabe rudranath er durotto kato? ’ or ‘What is the official distance of Rudranath from the Sea?’)

দৈত্যকুলাধিপতি অন্ধকের অত্যাচারে অতিষ্ঠ হয়ে দেবতারা কার শরণাপন্ন হন ? (‘Daityakuladhipoti andhak er otyachar e otistha hoe debotara kar sharonapanna hon?’ or ‘Whom did the gods take refuge being tortured by Andhak who was the king of daemons ? ’)

১৮৪২ অব্দে প্রকাশিত ফিশার্স কলোনিয়াল ম্যাগাজিনে দ্বারকানাথ ঠাকুর কে নিয়ে কি প্রকাশিত হয়েছিল? (‘1842 obde prokashito phisars colonial magazine e Dwaraka Nath Thakur ke nie ki prokashito hoechhilo?’ or ‘What was published in the Phishers’ Colonial magazine regarding Dwaraka Nath Tagore?’)

চিত্রাঙ্গদা নৃত্যনাট্যে কি কি চরিত্র আছে? (‘Chitrangoda nritynatye ki ki choritro achhe?’ or ‘What are the caharacters available in the dance-drama Chitrangada? ’)

The result is collected and accuracy, precision, F1 score are calculated. Accuracy of the result is compared with the other QA systems in Indian languages. F1 score is compared with the work in the SQUAD (Stanford Question Answer Dataset).

No new classification algorithm is invented in this work rather already existing simple classification techniques are applied for development question answering system in Bengali with the contribution to remove ambiguity and selection of the best classification technique among four. Though there are previous works for questions classification in Bengali but to the best of the knowledge there is no full phased working QA system in Bengali. So, the wok is novel in the ground of applying already existing algorithms in a new way to a new dataset.

2. Related Work

In Bengali fully working QA system is not available. Over time Researchers have tried to build different modules of the QA systems in Bengali. The significant and related research in this domain is presented below.

[1] have developed a closed domain factoid question answering system. The system has achieved 56.8% accuracy for answer extraction and 66.2% accuracy for returning the string containing the answer. It has achieved 90.6% accuracy for question classification.

Banerjee et al have worked exclusively in the question classification for Bengali QA system without going for answer retrieval [2]. They have demanded an accuracy of 87.79 % for question classification.

A Das and his group have applied statistical method for QA system in Bengali [3]. They have given rank to the probable answers and measured accuracy, precision and F1 score for appearance of the answer in the rank ranges. They have applied combination of seven statistical techniques to measure the score of an answer and summation of score for each answer is used to determine the rank.

[6] have used stochastic gradient descent (SGD) classifier for answer classification in the Bengali QA system. They have achieved an accuracy of 87.64% for fine grain classification and 95.56% for coarse grain classification.

[7] P P Manna and A Pal and have not gone for classification of the question for retrieving the answer. For a single match of the keyword they have retrieved the domains and after collecting all the domains they have ranked the domains according to the highest number of matches. After that the exact answer is retrieved from the best matched domain. They have achieved an overall 80% accuracy. The problem found in this work is that the exact answer may not be from best matched domain. In that scenario this system fails.

[8] have used classification techniques like Decision Tree, Naïve Bayes, Support Vector Machine and Stochastic Gradient Descent to classify the questions. They have achieved 90.5% accuracy over five categories of documents.

From 2001 the research on QA system showed significant growth. It started to include question classifier and answering agent started to work on classification of answers and information is retrieved from the specific class of the answer repository. Researchers started to work for improvement of the classifiers and most of them were targeted for English Question Answering System. Different Multinational Companies started to build QA system. IBM Watson is one such most popular and successful QA system. Stanford University created SQUAD dataset taking help of the crowd workers. The SQUAD [9] is the online available dataset in English where performance of any built QA system in English can be tested. Before 2001 also QA systems were existing. They were mostly domain centric. LUNAR [10] and BASEBALL [11] are the two early QA system. BASEBALL was made to give answers of the questions about baseball league. LUNAR was capable to give answers to the questions regarding geological data about rocks collected by Apollo Moon Mission.

Since inception, many researchers have tested the performance of their built system for SQUAD dataset. In recent times, Albert [12] achieved a very good performance. It showed 92.215 F1 score. Z Zhang and his group have developed retro reader [13,14] on top of the Albert system using ensemble method and achieved F1 score 92.978.

3. Problem Statement

Getting a query or question Q in natural language the system should

a) Determine the class C of the Q. That means, the system should categorize the question into any one of the eighty-six classes.
b) Return the answer string A. A should be the Exact Answer (EA) or A should contain EA.

4. Proposed Approach

Classification is conceptualized in this approach to predict the answer of a question. After receiving the question, preprocessing techniques are applied. Font similarization, detection of punctuation symbols, removal of extra space, and tokenization are applied in the first round of preprocessing. Next, the question sentence is passed to the LTRC shallow parser and the question sentence is received with all the words of the question are tagged with the parts of speech (POS). Using POS and grammatical rules, Function Words (FW), Content Words (CW), subjects and objects of the sentence, Named Entities and WH words are identified in the question sentence. Root form of the verb, person, tense and number are identified with the algorithm described in [5]. The same preprocessing method is applied to the sentences of the repository one by one.

In the next phase, using the classifier the question is classified into one of the eighty-six categories. Examples of the categories are – Anthropology, Astronomy, Biography, Chemistry, Child_Literature, Computer Science, Drawing, Economics, Geology, History etc.

Standard four classifiers are used for the classification. They are Naïve Bayes, Support Vector Machine, Decision Tree and Artificial Neural Network. Four classifiers are used to bring transparency and remove any biasness. If the result of four classification techniques are same, only one class is associated with the question. If four classification techniques derive more than one class of the question, all the classes are associated with the question. In the next phase, answer of the question is searched in the associated categories only.

After determining the class(es) of the question, search operation is concentrated into the associated class(es) only. The sentences of the repository of the determined class are classified recurrently upto reaching sentence level atomicity using the features FW, CW, person, tense, number, gender etc. to match the question sentence. At last the matched answer sentence is returned.

5. Algorithm

Input: The question from the user.
Output: The answer to the user.

Step 1: Parse the input question and tokenize.

Step 2: Pass the question to shallow parser (IIIT) and receive the tokens of the question sentence tagged with POS (Parts of Speech).

Step 3: Mark and collect Function Word, Content Word, subject and object of the sentence, Named Entity, Verb etc. with the help of POS and rules of grammar.

Step 4: Call “Das and Halder” algorithm [5] to extract the root form of the verb and identify person, tense, number of the subject of the sentence using the rules of grammar.

Step 5: Classify the question using the classifiers ANN, Naïve Bayes, SVM and Decision Tree based on attributes of the feature vector consisting Function Word, Content Word, Named Entity, root form of the verb, person, tense, number of the subject and WH word of the sentence (question).

Step 6: Repeat step 1 to step 5 for the sentences in the repository to reach the classification atomicity upto sentence level.

Step 7: Select the sentence with highest class and sub class match with the question in step 6.

Step 8: Try to form the exact answer (EA) from the selected sentence using rule-based knowledge base or return the selected sentence as A.

The detailed flowchart of the work is given in Fig. 3.

Extraction of Root Verb [5] is separate work, details of which is not described in the flowchart. The algorithm of Root Verb Extraction is given in the Fig.1.

Algorithm 1: DAS & HALDER

Input: Bengali corpus to Shallow parser (LTRC)
Output: Inflected verbs are collected in a file Input.doc
Input to our system: Input.doc
Output of our system: Multiple files classified according to tense and person consisting of root form of verbs

```
1  begin
2    Corpus is taken as input to the Shallow parser
3    All the inflected verbs are collected in a file named as Input.doc
4    Input.doc is taken as input to our system
5    if বিভক্তি(in the Table 1) matches then
6      if person matches then
7        Verbs are passed to the respective method with the verbs and বিভক্তি for processing
8        Verbs are processed one by one
9        Panini's rules are applied on those verbs to extract root verbs
10       Root verb is stored in Output.doc file as well as separate file according to tense.
11     else
12       The verb collected from Shallow parser in Input.doc is not a verb
13   else
14     The verb collected from Shallow parser in Input.doc is not a verb
```

Fig. 1. Automatic Root Verb Extraction

POS tagger is not developed in this work. Rather, LTRC Shallow Parser developed by the consortium of universities lead by IIT Bombay and hosted by IIIT Hyderabad is used for parts of speech tagging. So, the details of POS tagging are not discussed here.

Java is used for all interfacing and calling services or APIs. Weka classifier is used for classification. PostgreSQL relational database is used for knowledge base. Training sets are prepared without considering test sets. 10 folds cross validation is used throughout the experiment to balance the data set and remove any biasness.

Tense	Type of tense	Suffices (বিভক্তি)
বর্তমান কাল (Present Tense)	সামান্য বর্তমান Simple Present	"ি","ে","েন","িস", "ই"
	ঘটমান বর্তমান Present Continuous	"ছি","িতেছি","ছে","িতেছে","ছ","িতেছ","ছেন"
	পুরাঘটিত বর্তমান Present Perfect	"েছি","িয়াছি","েছ","িয়াছ", "েছে","েছেন"
অতীত কাল (Past Tense)	সামান্য অতীত Simple Past	"লাম","লুম","িলাম","িলুম","লে","িলে","লেন","িলেন"
	ঘটমান অতীত Past Continuous	"ছিলাম","ছিলুম","িতেছিলাম","িতেছিলুম"
	পুরাঘটিত অতীত Past Perfect	"েছিলাম","েছিলুম" ,"িয়াছিলাম"
	নিত্যবৃত্ত অতীত Past Perfect Continuous	"তাম","তুম","িতাম","িতুম" ,"তে"
ভবিষ্যৎ কাল (Future Tense)	সামান্য ভবিষ্যৎ Simple Future	"ব","িব","বে","িবে", "বি","িবি"
	ঘটমান ভবিষ্যৎ Future Continuous	"তেথাকব","িতেথাকিব",তেথাকবে","িতেথাকিবে","তেথাকবি","িতেথাকিবি",
	পুরাঘটিত ভবিষ্যৎ Future Perfect	"েথাকব","িয়াথাকিব", "েথাকবে"

Fig. 2. Table 1. Different kind of suffices applied to the verb in Bengali with Tenses

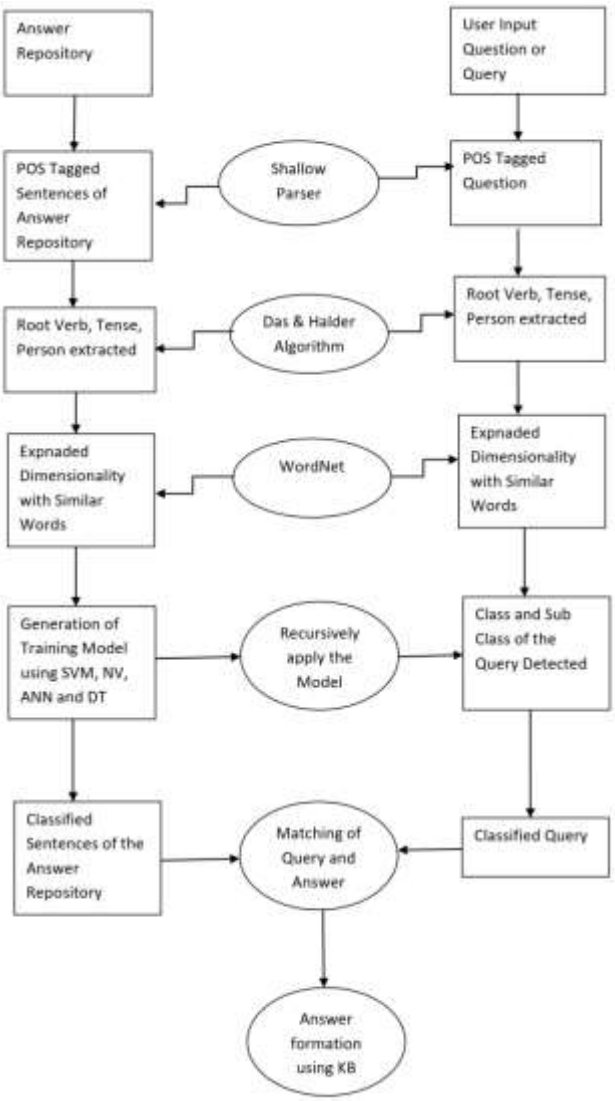

Fig. 3. Detailed Flowchart of the Methodology used

6. Result

6.1. *Result Summary*

The system detects the class of the question for 244 questions out of 250 questions accurately. Thus, for question classification, the method gives 97.6 percent accuracy.

In case of 234 questions out of 250 questions, the system returns the answer, A containing EA. Thus, coarse grain accuracy for answer prediction is 93.6 percent.

In case of 214 questions out of 250 questions, the system returns the answer A which is the EA.Thus, fine grain accuracy for answer prediction is 85.6 percent.

6.2. *Calculation of TP, TN, FP, FN, Accuracy, Precision, Recall and F1 score*

TP = True Positive = The predicted answer and the actual answer is same = 234
TN = True Negative = The repository does not have the answer of the question and the system predicts that the answer is not known = 5
FP = False Positive = The repository does not have the answer of the question but the system predicts an answer = 3
FN = False Negative = The repository has the answer of the question but the system predicts that the answer is not known = 8

Accuracy =(TP+TN)/ (TP+TN+FP+FN) = 95.6
Precision= TP/ (TP+FP) = 98.73
Recall = TP/(TP+FN) = 96.69
F1 score = 2*(Recall*Precision)/ (Recall + Precision) = 97.69

6.3. *Comparison of Result with other Methods and Dataset*

Accuracy percentage for question classification is compared in the Table 2.

Author	Language	Dataset	QC percentage
Monisha and her group	Bengali	SUST database	90.6%
Banerjee et al	Bengali	Own created dataset [2]	87.79%
M A Islam et al [6]	Bengali	University database	95.6%
Sarker et al	Bengali	SUST database	90.5%
Arijit Das	**Bengali**	**TDIL**	**97.6%**

Table 2. Comparison of Accuracy for Question Classification with Previous Work

F1 score percentage for the overall QA system is compared in the Table 3.

Author	Language	Dataset	F1 Score
Z Lan et al (Albert system)	English	SQuAD	92.215
Z Zhang et al	English	SQuAD	93.11
Arijit Das	**Bengali**	**TDIL**	**97.69**

Table 3. Comparison of F1 Score with Previous Work

This is never claimed that the performance of the presented system is more efficient than the systems tested in the SQuAD dataset. The reasons are stated below

i) SQuAD is a fully developed QA dataset. A million of Questions are available. Whereas in the presented work only 250 questions are

NV	True	False
Positive	212	20
Negative	8	10

tested.

Percentage of the accuracy is decreased with larger sets of question answer in general.

ii) In Table 3 all the experiments over SQuAD dataset are

SVM	True	False
Positive	210	25
Negative	5	10

implemented using deep learning. Whereas the presented work has used fully supervised method. The performance of supervised method is always better than

ANN	True	False
Positive	233	3
Negative	7	7

hybrid method in general.

iii) As the dataset structure is different, TDIL and SQuAD should not be compared but there is no standard dataset in Bengali where this work can be compared.

Presenting Table 3 was not to show that the presented system is better than the systems designed by the Shanghai Jiao Tong University, but it proves that the carried-out research work is in the right direction and the work has its own value.

6.4. *Detailed Result for Individual Classifier*

Confusion matrix summarizes prediction result with respect to the actual result. True Positive (TP) is the case where the predicted class and the actual class are same. True Negative (TN) is the case when theactual answer does not exist in some class or subclass and prediction also tells about that absence. False Positive (FP) is the case when the actual answer does not exist in some class or subclass but the prediction tells that answer is contained in that class or subclass. False Negative (FN) is the case when actual answer is contained in the class or subclass but the prediction tells that the answer is not present.

Table 4, 5, 6, 7 depicts respectively the confusion matrices for Naïve Bayes, SVM, ANN and Decision Tree. Out of 250 questions, what is the value for TP, TN, FP and FN are shown in the tables for all the four classification techniques.

Table 4. Confusion Matrix (Naïve Bayes)

Table 5. Confusion Matrix (SVM)

Table 6. Confusion Matrix (ANN)

DT	True	False
Positive	179	33
Negative	3	35

Table 7. Confusion Matrix (Decision Tree)

7. Evaluation of Result

In case of classification and sub-classification ANN gives the most accurate result. In the Table 6. Accuracy, Precision, Recall, F1 Score are shown for all the four algorithms.

	Accuracy	Precision	Recall	F1 Score
Formula	(TP+TN)/(TP+TN+FP+FN)	TP/(TP+FP)	TP/(TP+FN)	2*(Recall*Precision)/(Recall + Precision)
NV	88	91.37	95.49	93.38
SVM	86	89.36	95.45	92.30
ANN	96	98.72	97.08	97.89
DT	72.8	84.43	83.64	84.03

Table 8. Calculation of Accuracy, Precision, Recall and F1 Score

8. Analysis of Error

It is observed in case of question classification that, when the question appears to be from more than one class the classifier gets confused. For example, the question "স্মৃতি ইরানি কে?" ('Smriti Irani Ke?' or 'Who is Smriti Irani') is classified as "Entertainment" category by SVM and Decision Tree and as "Politics" category by ANN and Naïve Bayes. The actual answer (EA) was available in the "Politics" category of text.

In case of answer retrieval, if the answer is expanded in more than one sentence or for the question which are not factoid the classifier often fails to link those sentences.

9. Few Closed Observations

a) When any question is received by the system for first time the response time is too long. This is because of high run time of classification algorithms. If the questions are repeated the response time is less. Though, the answer of any question is not saved but for repetition of the question re-classification is not required as the model is already available.

b) When the sentences are linked with anaphora, the system does not perform perfectly to retrieve the answer. The classifier gets confused to classify different linked sentences.

c) The system performs degrades with the use of foreign words. This happens when there is any English or Hindi mixed word in the question. The word is not found in the WordNet so it causes problem.

d) The system performs as expected when the language of the training set and the language of the testing set are same. The system performance degrades for code-mixed language when training is not done with the code-mixed language.

10. Application

QA system is an intersection of NLP and IR. To find the exact answer or the string containing the answer, QA system is used. It is different from search engine where list of URLs is returned rank wise. QA system can be implemented in audio medium or text medium. In case of audio medium, a voice to text algorithm is used.

QA system is used for chat bot, virtual assistants etc. Examples of virtual assistants are Siri, Alexa, Cortana, Google Assistant etc.

QA system has a direct application in the automatic response system (ARS). ARS is implemented in different industry to reduce the work load on call center. QA system is easily found in banking, auto mobile, health, tele medicine industries. More and more industries are lining up to limit the burden of excess workload in customer care.

It has a huge importance in Automatic Mail Reply, Story Telling, Music Synthesizer, Sentiments Analysis etc.

11. Scope for Improvements

The QA system is built on supervised learning algorithms. Huge time and labour is required to make the training dataset for supervised method. Though unsupervised and statistical procedures are applied in [4], it reduces the percentage of accuracy. Thus, possible way of improvement is to use the hybrid approach where labour and time required for training will be less but accuracy of the result will be high.

Incorporating human feedback for the test result in the training during next iteration is another possible improvement for increasing accuracy.

12. Conclusions

In this work, QA system in Bengali is developed with the help of four major supervised learning methods. Getting the question, the system classifies it into one category with the help of classification algorithm. Then particular that category is searched for returning answer. If more than one category is determined by the four classification algorithms then all the categories are searched.

Here, searching is not done syntactically or based on key word matching. Rather, classification is used recursively to reach the sentence level of atomicity for returning the answer string.

Recent trends of research are to employ the deep learning which has been discussed in the 'Related Work' section on the SQUAD dataset. Deep learning needs costly hardware resources. A series of GPU workstations or cloud-based services are required to train the system efficiently. This requires the university or research center to invest a huge amount to build such facility or laboratories.

On the contrary the method discussed in this paper is light-weight and can run on personal computers or laptops. Though the described method needs higher manpower to build the training set but this supersedes on the point of hardware cost involved in deep learning as both requires almost same time. This is another ground to describe this work as novel in such a country where huge manpower is available. This is extremely relevant in this COVID-19 pandemic period when there is scarcity of fund to build new laboratory or facility.

13. Declaration

11.1 *Availability of Data and Material*

The data that was used to do this experiment is available with Language Research Unit of ISI Kolkata. It is a product of TDIL (Technology Development of Indian Language) project of Ministry of Electronics and IT, Govt. of India. For the experiment the access of this data is received from Prof. N S Dash, Professor and Linguist of ISI Kolkata. The data may be shared upon reasonable request from academicians or research scholars.

11.2 *Competing Interests*

The author declares that he has no competing interests.

11.3 *Funding*

There is no funding received directly for this work from any organization. Arijit Das was supported by the UGC UPE fellowship when he was carrying out this work.

13.4. *Acknowledgements*

LTRC group, IIIT Hyderabad has provided Shallow Parser for POS tagging. Professor Niladri Sekhar Das of Indian Statistical Institute, Kolkata (ISI Kolkata) has made the dataset available. He being the renowned linguist also helped in evaluation.

References

1. Sarker, S., Monisha, S.T.A. and Nahid, M.M.H., 2019, September. Bengali Question Answering System for Factoid Questions: A statistical approach. In *International Conference on Bangla Speech and Language Processing (ICBSLP)* (Vol. 27, p. 28). DOI: 10.1109/ICBSLP47725.2019.201512.
2. Banerjee, S., Naskar, S.K., Rosso, P. and Bndyopadhyay, S., 2019. Classifier combination approach for question classification for Bengali question answering system. *Sadhana*, *44*(12). DOI: 10.1007/s12046-019-1224-8.
3. Das, A., Mandal, J., Danial, Z., Pal, A.R. and Saha, D., 2019. A Novel Approach for Automatic Bengali Question Answering System using Semantic Similarity Analysis. *arXiv preprint arXiv:1910.10758*.
4. Arijit Das and DigantaSaha, "Improvement of electronic governance and mobile governance in multilingual countries with digital etymology using sanskrit grammar," in Proc. IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 2017, pp. 502 – 505. DOI: 10.1109/R10-HTC.2017.8289008.
5. Arijit Das, Tapas Halder and DigantaSaha, "Automatic extraction of Bengali root verbs using Paninian grammar" in Proc. 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), Bangalore, India, 2017, pp. 953 – 956. DOI: 10.1109/RTEICT.2017.8256739.
6. Islam, M.A., Kabir, M.F., Abdullah-Al-Mamun, K. and Huda, M.N., 2016, May. Word/phrase based answer type classification for bengali question answering system. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)* (pp. 445-448). IEEE. DOI: 10.1109/ICIEV.2016.7760043.
7. Manna, P.P. and Pal, A.R., 2019, May. Question Answering System in Bengali Using Semantic Search. In *2019 International Conference on Applied Machine Learning (ICAML)* (pp. 175-179). IEEE. DOI: 10.1109/ICAML48257.2019.00041.
8. Monisha, S.T.A., Sarker, S. and Nahid, M.M.H., 2019, May. Classification of bengali questions towards a factoid question answering system. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-5). IEEE. DOI: 10.1109/ICASERT.2019.8934567.
9. Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P., 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*.
10. Woods, W.A., 1973, June. Progress in natural language understanding: an application to lunar geology. In *Proceedings of the June 4-8, 1973, national computer conference and exposition* (pp. 441-450). DOI: 10.1145/1499586.1499695.
11. Green Jr, B.F., Wolf, A.K., Chomsky, C. and Laughery, K., 1961, May. Baseball: an automatic question-answerer. In *Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference* (pp. 219-224). DOI: 10.1145/1460690.1460714.
12. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R., 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.
13. Zhang, Z., Yang, J. and Zhao, H., 2020. Retrospective reader for machine reading comprehension. *arXiv preprint arXiv:2001.09694*.
14. Zhang, Z., Yang, J. and Zhao, H., 2020. Retrospective reader for machine reading comprehension. *arXiv preprint arXiv:2001.09694*.
15. A Das and D Saha, "**A Novel Approach to Enhance the Performance of Semantic Search in Bengali usingNeural Net and other Classification Techniques,**" in arXiv:1911.01256.