**Smart improvised Question Answering System using BERT, LSTM and Semantic Structuring based on University’s dataset.**

**Bimal P. a, Aman S.a, Yashas M. a, Arjun K. a, Neha P. a, Saravanakumar K. a**

a: *School of Computer Science and Technology, Vellore Institute of Technology, Vellore,, Tamil Nadu 632014, India*

# Abstract:

Every large organization has a massive collection of documents that it can use to build a Question Answering systems to serve and solve the queries of stakeholders in that institute. Hence we propose a Question Answering System based on official documents available in a university. Although multiple solutions have existed for Question Answering systems. We are proposing the use of stacked layer of LSTMs (Long short term memory) to get the corresponding answer for the particular question where the words (tokens) are fed as input to the neural networks. We are using the BERT model for keywords extraction and feeding it to the LSTM neural net. This solutions is being built especially useful for Universities that have very large number of Normative and Directive documents. This solution can be used to quickly provide response to students’ queries whenever they visit the site without requiring them to visit individual sub components of the site. It makes the use of existing dataset in the organization while saving a lot of time for the students and staffs.

# Introduction:

Question answering is the branch of computer science that deals with building automatic systems that can provide answers to human generated Natural Language questions [79]. It is the blend of Natural Language Processing with Machine Learning techniques and Semantic Analysis [5]. Question answering system have come across a long way through since its initial inception in the early 1960s [80]. Initially, it used to be a simple single domain system trained on limited amount of data E.g.: LUNAR [80], BASEBALL [81], etc. The Gradually, as the sizes of knowledge bases(KB) grew larger and larger, more complex systems started to be developed that made use of more computation to process massive amount of data to develop more smarter QA systems[82]. First, there used to be linear models based Machine learning approaches like Support vector machines and Logistic regression and based on spare vectors of very high dimensions to solve the problems in Natural Processing domain [83]. Gradually, as technologies progressed and hardware became more capable, QA systems slowly progressed to non-linear models based on graphs, trees and neural networks.

On the basis of domain/scope of topics that the QA model addresses and responds to is divided into 3 major types: Closed domain QA and Open Domain QA. Close domain Question answering, as name implies, mostly have a narrow scope and are simpler while the open domain systems are comparatively more complex and require massive amount of data for training it [84].

A typical "Question Answering" model based on modern NLP techniques involves the following 3 broad steps while working on production: Question Preprocessing and Classification, Document Searching and Information retrieval, Answer Extraction, analysis and sentence generation [82].

In modern era, there are a lot of large organizations that have huge amount of structured and unstructured data in digital forms. Most of them are available publicly in the websites of the respective organizations while the others are internally maintained. These documents can range from norms, values of the organizations, code of conducts, rules, regulations and directions, etc. These documents can be a huge source of knowledge and information that we can use to build organization specific question answering system such that it resolved the queries of the students of the respective institutions. Building Question Answering Systems out of these is the most promising way to use and retrieve data from such knowledge bases by the end user [5].

It is a common problem faced by a lot of students that requires information are available on the university website but the students find it tedious to go through the site and into specific sub segment and get answer to their queries. It also takes a lot of their time and effort. Numerous solutions exist in this domain that are built for general purpose question answering but very few solutions exist that actually are made for specific type of documents like the strictly formal documents that organizations such as an university might have [7]. So, we are proposing a different model of question answering that specifically solves the problem in University and helps the students, staffs or anyone who might have some queries related to the institutions whose answers maybe be present in their documents or webpages.

For the solution to the chosen problem, we are using specially customized LSTMs with semantic structuring and BERT models for extracting the keywords. The extracted keywords will be sent to the customized LSTM layers and further processed to generate most probable and accurate answers to the queries passed. After the LSTM generated the most probable keywords in the answers to the questions, we further apply lemmatization and semantic structuring to form a sensible sentence using the keywords to give back reply to the user.

# Related Works:

For question answering (QA), to achieve good results, current state-of-the-art pre-trained models often require fine-tuning on tens of thousands of samples [1]. In a few-shot setting, their performance suffers dramatically (< 100 examples). A simple fine-tuning framework that employs pre-trained text-to-text models and is closely aligned with their pre-training framework was one of the offered solutions to address this [1]. The suggested few-shot fine-tuning framework design differs from the current norm for QA fine-tuning frameworks in terms of input-output design and training purpose [1].

Question Answering (QA) is the task of enabling a machine to automatically answer questions posted by humans in a natural language form [2]. Community question answering is the process of selecting the best answer from a pool of candidate answers (cQA) [2]. They develop a cross-sentence context-aware bi-directional LSTM model that generates hidden representations for both the question and answer texts, allowing them to be aware of each other's context [2].

[7] is making use of normative documents (i.e., formal documents that contain strict formatting and formal languages such as legal documents, agreements, rulebooks, affidavits, etc.) to build Question Answering System by performing interactive QA for highly precise results.[7][47] Reading such documents is hard. The name of the approach is Interactive Question Answering using pre-processed normative documents and algorithm used is Semgrex (Special version of regular expressions) [49]. So, they are building a chatbot is a kind of question answering (QA) system that enables answering users’ questions automatically.[48] To use the open domain systems in closed domain applications. [50]

[8] They are finding solution to train a Natural language-based Question Answering Model effectively and functionally in constrained scenarios where there is very less amount of training data [51][52]. The name of approach is Pseudo labelling with DistilBERT (transformers) for training QA systems with less data and algorithm used is Semi supervised learning-based pseudo labelling technique [53][54]. For that they trained a deep neural network in supervised manner simultaneously with labelled and unlabeled data [55].

[3] They have extend the boundaries of E2E learning for KGQA to include the training of an ER component as there are challenging problems in other datasets [56]. The proposed solution is approached as KGQA rely on semantic parsing (SP) to translate natural language into a logical form and algorithm used is RL (reinforcement learning) [57][58]. When compared with [59] they cannot be answered due to context ambiguity.

The paper [4] deals with the problem that creating datasets for Question Answers (QA) requires expensive hand-labelling work and specialized feature engineering. The focus of this paper is to collect more and better data that can maximizes model potentials [60][61]. The proposed solution describes their experimental setup, datasets, and then applies RWS to AS2-NQ and algorithm used is RoBERTa [62] and ELECTRA [63].

The objective of [10] in this paper was to provide a more effective data organization, management, and cognitive ability for the customer service business which other papers can’t do. [64][65] Therefore the proposed solution for this was Ontology construction and Construction of Knowledge Graph in the Field of Power Customer Service, algorithm used for this was BiLSTM-CRF. [66][67][68] Datasets used were 95598 electric power customer service business data. [69]

[15] Their research addresses numerous elements such as model size, quality of pretrained models, amount of data synthesised, and algorithmic decisions in order to close this gap using large language models [70][71] The proposed approach is a three-step pipeline that includes unconditional answer extraction, question generation, and question filtration, with BERT, GPT-2, and SQUAD as the algorithms. [72][73] Some of the datasets used are complex web questions. [74]

In the paper [13], they introduce the Sim2RealQA framework, which thoroughly trains a QA model with QA datasets generated in a life simulator and then uses it to solve real-world QA problems without answer labels. Algorithm used is RNN, DMN [77][78].

To perform improvised Question Answering by building and querying SPARQL query templates [5]. It achieves higher efficiency than existing systems and makes good uses of large amounts of semantic data [5] available over the web. The algorithm used to issue this problem is Query Formulation algorithm in [5]. The solution is composed of two vital sections: Query processor [5] and Sparql query builder [5].

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SQA Processes** | **Paper ID** | | | | | | | | | |
|  | **17** | **20** | **18** | **19** | **8** | **6** | **7** | **5** | **9** | **12** |
| **Process 1: Capture NL query** |  | | | | | | | | | |
| Simple question/non question form | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Process 2: Syntactic analysis** |  | | | | | | | | | |
| Part of speech tagging | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |
| POS tagger and dependency parser | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Stanford parser | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Identify noun phrase/noun concept | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ |
| Named Entity Recognition | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Morphological analysis | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |  | ✗ | ✗ |
| Lemmatization | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Tokenization | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Question Analysis | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ |
| Question type classification | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ |
| Syntactic sentence analysis | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| **Process 3: semantic analysis** |  | | | | | | | | | |
| Synonym | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Homonym | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Hypernym | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| **Process 4: Syntactic and semantic linguistic disambiguation** | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ |

# Proposed Approach:

We have proposed the following model for the improvised Architecture model for the Question Answering based on BERT and stacked layers of LSTM, semantic and semantic structuring to serve the best answer corresponding to the given query.

## 3.1 Design of Architecture:

This research proposes a generic tool for decision-makers that would help the students of a particular university get answers to the common related doubts and question (FAQ) to their university. This could be an add-on (as a widget/ bot) to various university websites where this question-answering model can help to answer a student’s doubt almost instantly. This model is trained on a university dataset (self-created for this task) along with various popular QA datasets. Coming to the architecture this model is novel from any other existing architecture as this uses “keyword extraction” along with semantic restructuring to generate an answer to a particular question. You can see the proposed architecture below which uses BERT model to generate important keywords, feed them into a stacked LSTM network and then using semantic structuring to generate an appropriate answer to a question almost instantly.

Since there were no sufficient previously available datasets for our chosen problem domain, we created our own dataset to perform the experimentation and model testing. The dataset consisted of handpicked question-answer pairs extracted from the VIT University websites that included all different types of questions like the factoid Questions, list type Questions, casual type questions, confirmation type (Yes/No and True/False) Type Questions. The model was initially trained on a wide variety of datasets that included data from large datasets of formal style like the SQuAD from Wikipedia [86], Natural Questions (NQ) by Google, Question Answering in Context (QuAC), Conversational Question Answering (Coca) [87], HOTPOTQA [88], etc. Although most questions asked in the University’s FAQs will be closed domain (related to the University, administration, academics and system) factoid style (consisting of single answer and factual in nature, starting with [what, when, which, who.. etc.])[85], we made the model capable enough to handle the non-factoid questions as well just in case any such queries is put forward by the user while the model is running.

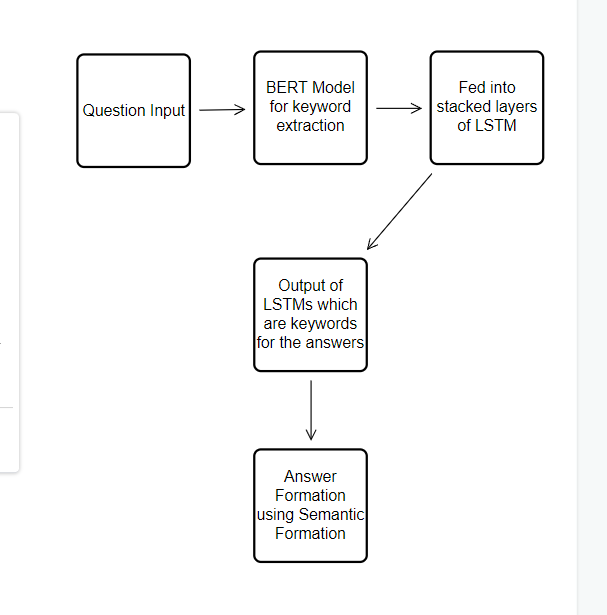


Fig 1: Simplified flow of the proposed model for Question Answering

## Datasets used:

We have used the following datasets for training our model:

* [SQuAD](https://rajpurkar.github.io/SQuAD-explorer/) from WikiPedia
* Natural Questions (NQ) by Google
* Question Answering in Context (QuAC)
* Conversational Question Answering (Coqa)
* HOTPOTQA

And for further testing it, we are also using a custom made dataset prepared by scraping data available on the university website. We are using BS4 and Scrapy libraries in Python to scrape the data and prepare it for testing the model.

# Experimentation:

The implementation of the proposed architecture consists of a BERT model to extract keywords (from the questions), then the use of LSTMs along with semantic structuring to generate the answer based on the output keywords from the LSTM (stacked neural network) We have trained and done the testing of the model using the university synthetic dataset which we have created to check its response and accuracy (metric like BLEU score and F1 score).

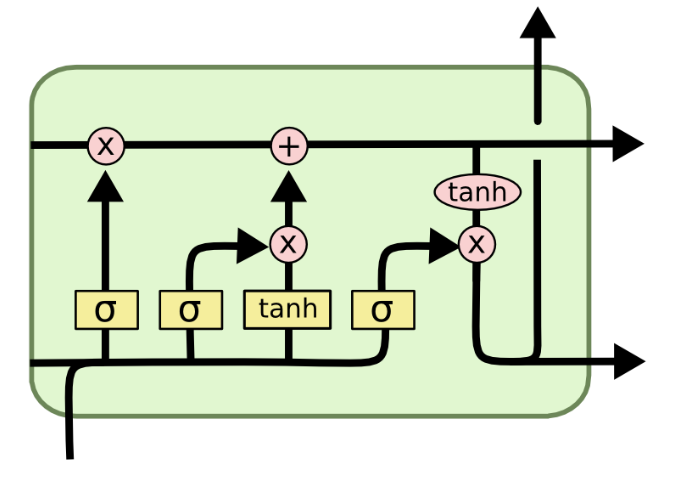
## 4.1 TRAINING

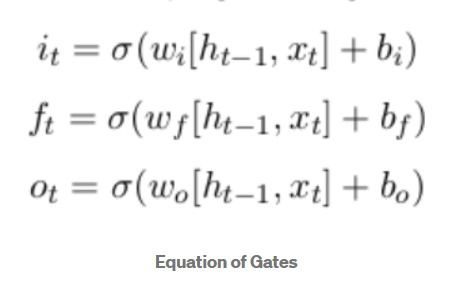
We train the model using the binary cross-entropy loss function:

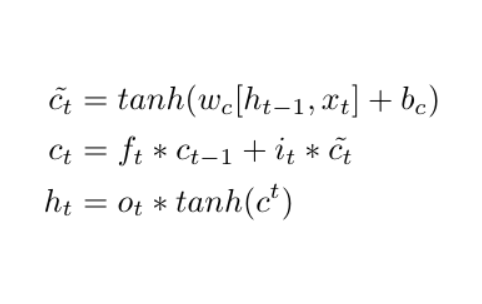
L(y, ˆy) = 1 NE X NE i=1 yi log ˆyi+ (1−yi) log (1−yˆi),

Where y ∈ R NE is a k-hot label vector.

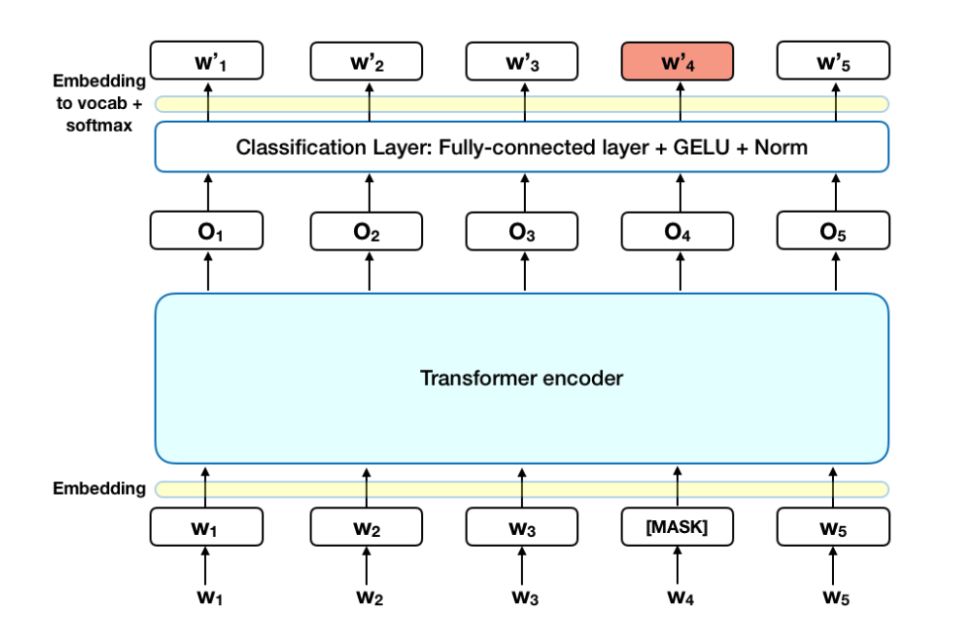
Apart from using BERT and semantic structuring, stacked layers of LSTMs are used to compute the keywords which should be present in the answer. Rather than a bidirectional LSTM model a Unidirectional model is used which uses the concept of backpropagation through time ( using gradient descent ) and binary cross entropy loss function defined above. Some of the equations of a LSTM layer to calculate the outputs during forward propagation are:







Standard BERT Model [50]:



The name of the model we have trained is called ASKE (Answer Segmentation using Keyword Extraction). Our current architecture currently is made up of three main parts which are the BERT model for keyword extraction, a series of stacked LSTM to output the necessary keywords to be present in the answer and then using semantic structuring to properly output the complete answer. Our architecture is a novel idea which gives us amazing results after the model is trained properly. To test our model we tested this on a university dataset . We use Adam optimizer with a learning rate of 2e5. We use a training batch size of 4. We don’t use learning rate scheduling.

## 4.2 Testing and Evaluation

To test our model infrastructure we have created a synthetic university dataset and used the training examples there to check the accuracy of our model (using the F1 score and BLEU score metric ). This gave us pretty amazing results on the dataset thus acting as a validation for our model. We also used word embedding to represent keywords generated from the BERT model which is then fed to the stacked layers of LSTM on top of each other. Each keyword generated from the question is fed to the neural network layers as an embedding matrix which is used to generate the keywords that should necessarily be present in the air. Below you will see the outputs generated by our model on the university synthetic dataset which we created and we can observer the model outputs are really amazing.

|  |  |
| --- | --- |
| Input | Output |
| 1. What is the nirf ranking of VIT Vellore? | 21 overall in India and 13 among Universities |
| 2. Where is VIT Vellore located? | VIT, Vellore Campus, Tiruvalam Rd, Katpadi, Vellore, Tamil Nadu 632014 |
| 3. How to get into VIT as an undergrad student? | Students must give an exam known as VITEEE |
| 4. What is the boy girl ratio in VIT? | Male to female percentage is found to be 74% : 26% in VIT-Vellore Institute of Technology |
| 5. What are chapters in VIT? | Chapters form for a variety of reasons such as: electing a central voice for issues that affect the entire class, to raise funds and to build a sense of camaraderie among classmates. |
| 6. What are the different types of mess available in VIT? | Special mess, Non-veg mess and veg mess |
| 7. Does VIT have a library? | Yes, it was named as Periyar Evar Central Library |

## 4.3 Comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **SQuAD** | **TriviaQA** | **NewsQA** | **HotpotQA** |
| **BART** | **10.44±5.9** | **2.5±2.1** | **2.9±1.4** | **5.6±2.4** |
| **ASKE** | **10.52±5.6** | **2.4±1.8** | **2.7±1.2** | **5.8±1.9** |

*Comparison of F1 scores across all datasets for the standard QA fine-tuning objective (BART) vs the proposed aligned fine-tuning objective (). The value after ± indicates the standard deviation across 5 runs with different seeds. NQ stands for Natural Questions. TbQA stands for TextbookQA.*

# Conclusions:

The paper’s aim is accomplished and a practical, novel approach to build a Question Answering system based on University’s data available in their websites have been examined. The proposed approach made use of BERT model for keyword extraction and then stack of unidirectional LSTM based neural network for answer keyword extraction which was further converted to readable answer using semantic restructuring techniques. The proposed methodology has also shown really good results and stand against the major solutions for similar problems in open domain QA available today. This also helped us to make a good use of the university’s data- that is just stored on the websites to build a functional QA system for the convenience of students/faculties/staffs in the university.

# Future research - discussion:

Although the proposed system was functional and delivered answers to the user queries, it still has a lot of improvements and optimizations. We propose the following

* Fine tuning with even larger data.
  + The amount of data that we used for training and testing from the University’s website was simply minimal and not that enough. So there was some problem seen in the fine tuning of the model. So, as a further work of this paper, we have kept the fine tuning of the obtained model using even larger amount of data by scraping through entire website. It would lead to better and more accurate
* Implementing the model in the actual website.
  + The purpose of any research is to build actual product and make lives easier. Adhering to this principle, we have planned to implement the model as a FAQ answering system in the real website and serve the students in the university.

# Citations:

[1] Chada, R., & Natarajan, P. (2021). FewshotQA: A simple framework for few-shot learning of question answering tasks using pre-trained text-to-text models. arXiv preprint arXiv:2109.01951.

[2] Wu, J., Mu, T., Thiyagalingam, J., & Goulermas, J. Y. (2020). Building interactive sentence-aware representation based on generative language model for community question answering. Neurocomputing, 389, 93-107

[3] Oliya, A., Saffari, A., Sen, P., & Ayoola, T. (2021). End-to-End Entity Resolution and Question Answering Using Differentiable Knowledge Graphs. arXiv preprint arXiv:2109.05817.

[4] Krishnamurthy, V., Vu, T., & Moschitti, A. (2021). Reference-based weak supervision for answer sentence selection using web data. arXiv preprint arXiv:2104.08943.

[5] Dhandapani, A., & Vadivel, V. (2021). Question Answering System Over Semantic Web. IEEE Access, 9, 46900-46910

[6] Manjunath, T. N., Yogish, D., Mahalakshmi, S., & Yogish, H. K. (2021). Smart question answering system using vectorization approach and statistical scoring method. Materials Today: Proceedings

[7] Pisařovic, I., Dařena, F., Procházka, D., & Janiš, V. (2022). Preprocessing of normative documents for interactive question answering. Expert Systems with Applications, 191, 116314.

[8] Kuligowska, K., & Kowalczuk, B. (2021). Pseudo-labeling with transformers for improving Question Answering systems. Procedia Computer Science, 192, 1162-1169

[9] Zhou, L., Small, K., Zhang, Y., & Atluri, S. (2021). Generating Self-Contained and Summary-Centric Question Answer Pairs via Differentiable Reward Imitation Learning. arXiv preprint arXiv:2109.04689

[10] Palumbo E ， Rizzo G ， Troncy R ． Entity2rec ： Learning user-item relatedness from knowledge graphs for top-n item recommendation [C] // Proceedings of the 11th ACM Conference on Recommender Systems．Como，Italy： ACM，2017：32-36．

[11] Zhong, B., He, W., Huang, Z., Love, P. E., Tang, J., & Luo, H. (2020). A building regulation question answering system: A deep learning methodology. Advanced Engineering Informatics, 46, 101195.

[12] Wu, W., Deng, Y., Liang, Y., & Lei, K. (2020). Answer Category-Aware Answer Selection for Question Answering. IEEE Access, 9, 126357-126365.

[13] Puri, R., Spring, R., Patwary, M., Shoeybi, M., & Catanzaro, B. (2020). Training question answering models from synthetic data. arXiv preprint arXiv:2002.09599.

[14] Miyanishi, T., Maekawa, T., & Kawanabe, M. (2021). Sim2realqa: Using life simulation to solve question answering real-world events. IEEE Access, 9, 75003-75020.

[15] Dasigi, P., Lo, K., Beltagy, I., Cohan, A., Smith, N. A., & Gardner, M. (2021). A dataset of information-seeking questions and answers anchored in research papers. arXiv preprint arXiv:2105.03011.

[16] Yu, J., Zhu, Z., Wang, Y., Zhang, W., Hu, Y., & Tan, J. (2020). Crossmodal knowledge reasoning for knowledge-based visual question answering. Pattern Recognition, 108, 107563.

[17] Xiong, P., Zhan, H., Wang, X., Sinha, B., & Wu, Y. (2019). Visual query answering by entity-attribute graph matching and reasoning. In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition (pp. 8357-8366).

[18] Cao, Q., Liang, X., Li, B., & Lin, L. (2019). Interpretable visual question answering by reasoning on dependency trees. IEEE transactions on pattern analysis and machine intelligence, 43(3), 887-901.

[19] Xu, B., & Zhuge, H. (2020). The influence of semantic link network on the ability of question-answering system. Future Generation Computer Systems, 108, 1-14.

[20] Barskar, R., Ahmed, G. F., & Barskar, N. (2012). An approach for extracting exact answers to Question Answering (QA) system for english sentences. *Procedia Engineering*, *30*, 1187-1194.

[21] Alberti, C., Andor, D., Pitler, E., Devlin, J., & Collins, M. (2019). Synthetic QA corpora generation with roundtrip consistency. arXiv preprint arXiv:1906.05416.

[22] Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., & Deng, L. (2016, January). MS MARCO: A human generated machine reading comprehension dataset. In CoCo@ NIPS.

[23] Du, X., & Cardie, C. (2017, September). Identifying where to focus in reading comprehension for neural question generation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2067-2073).

[24] Du, X., & Cardie, C. (2018). Harvesting paragraph-level questionanswer pairs from wikipedia. arXiv preprint arXiv:1805.05942.

[25] He, J., Kryściński, W., McCann, B., Rajani, N., & Xiong, C. (2020). Ctrlsum: Towards generic controllable text summarization. arXiv preprint arXiv:2012.04281.

[26] Rao Zhuyi Zhang Yunxiang．Research on intelligent customer service system based on knowledge map [J]．Electric Power Information and Communication Technology，2017，15(7)：41- 45(in Chinese)．

[27] Cheng S ， Shen J ， Shi Q ， et al ． Research on the construction of three level customer service knowledge graph [C] // 2017 3rd International Conference on Applied Materials and Manufacturing Technology．Changsha，China：ICAMMT，2017：1-6．

[28] Liang Linsen ． Research on innovative application of customer service robot in power supply business hall based on Intelligent Q&A technology [J]．Wireless Internet Technology，2019，16(4)：139- 140(in Chinese)．

[29] Zhou fan，Ye Jianhui，Xiao Linpeng，et al．Research on Intelligent Question Answering System of power grid model ontology

[30] Cheng, C. P., Lau, G. T., Law, K. H., Pan, J., & Jones, A. (2009). Improving access to and understanding of regulations through taxonomies. Government Information Quarterly, 26(2), 238-245

[31] Zhong, B. T., Ding, L. Y., Luo, H. B., Zhou, Y., Hu, Y. Z., & Hu, H. M. (2012). Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking. Automation in Construction, 28, 58-70.

[32] Yang, Y., Yih, W. T., & Meek, C. (2015, September). Wikiqa: A challenge dataset for open-domain question answering. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 2013-2018).

[33] Sahu, T. P., Nagwani, N. K., & Verma, S. (2016). Multivariate beta mixture model for automatic identification of topical authoritative users in community question answering sites. IEEE Access, 4, 5343- 5355.

[34] Bian, W., Li, S., Yang, Z., Chen, G., & Lin, Z. (2017, November). A compare-aggregate model with dynamic-clip attention for answer selection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 1987-1990).

[35] Chen, Q., Hu, Q., Huang, J. X., He, L., & An, W. (2017, August). Enhancing recurrent neural networks with positional attention for question answering. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 993-996).

[36] Rao, J., Liu, L., Tay, Y., Yang, W., Shi, P., & Lin, J. (2019, November). Bridging the gap between relevance matching and semantic matching for short text similarity modeling. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 5370-5381).

[37] Heilman, M., & Smith, N. A. (2010, June). Good question! statistical ranking for question generation. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 609-617).

[38] Joshi, M., Choi, E., Weld, D. S., & Zettlemoyer, L. (2017). Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. arXiv preprint arXiv:1705.03551.

[39] Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019). Ctrl: A conditional transformer language model for controllable generation. arXiv preprint arXiv:1909.05858

[40] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

[41] Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3), 542-542.

[42] Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A., & Starner, T. (2004, April). Recognizing workshop activity using body worn microphones and accelerometers. In International conference on pervasive computing (pp. 18-32). Springer, Berlin, Heidelberg.

[43] Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y., & Okadome, T. (2010, May). Object-based activity recognition with heterogeneous sensors on wrist. In International Conference on Pervasive Computing (pp. 246-264). Springer, Berlin, Heidelberg.

[44] Blum, M., Pentland, A., & Troster, G. (2006). Insense: Interest-based life logging. IEEE MultiMedia, 13(4), 40-48.

[45] Lester, J., Choudhury, T., & Borriello, G. (2006, May). A practical approach to recognizing physical activities. In International conference on pervasive computing (pp. 1-16). Springer, Berlin, Heidelberg.

[46]: Bewley, A., Rigley, J., Liu, Y., Hawke, J., Shen, R., Lam, V. D., & Kendall, A. (2019, May). Learning to drive from simulation without real world labels. In 2019 International conference on robotics and automation (ICRA) (pp. 4818-4824). IEEE.

[47] Camilleri, J. J., Haghshenas, M. R., & Schneider, G. (2018, April). A web-based tool for analysing normative documents in English. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 1865-1872).

[48] Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017, July). Superagent: A customer service chatbot for e-commerce websites. In *Proceedings of ACL 2017, system demonstrations* (pp. 97-102).

[49] Ryu, P. M., Jang, M. G., & Kim, H. K. (2014). Open domain question answering using Wikipedia-based knowledge model. *Information Processing & Management*, *50*(5), 683-692.

[50] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

[51] Bjerva, J., Bhutani, N., Golshan, B., Tan, W. C., & Augenstein, I. (2020). SubjQA: a dataset for subjectivity and review comprehension. *arXiv preprint arXiv:2004.14283*.

[52] Muttenthaler, L. (2020). Subjective Question Answering: Deciphering the inner workings of Transformers in the realm of subjectivity. *arXiv preprint arXiv:2006.08342*.

[53] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

[54] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, *2*(7).

[55] Lee, D. H. (2013, June). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML (Vol. 3, No. 2, p. 896).

[56] Dahl, D. A., Bates, M., Brown, M. K., Fisher, W. M., Hunicke-Smith, K., Pallett, D. S., ... & Shriberg, E. (1994). Expanding the scope of the ATIS task: The ATIS-3 corpus. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*.

[57] Sun, H., Arnold, A., Bedrax Weiss, T., Pereira, F., & Cohen, W. W. (2020). Faithful embeddings for knowledge base queries. *Advances in Neural Information Processing Systems*, *33*, 22505-22516.

[58] Wu, P., Huang, S., Weng, R., Zheng, Z., Zhang, J., Yan, X., & Chen, J. (2019). Learning representation mapping for relation detection in knowledge base question answering. *arXiv preprint arXiv:1907.07328*.

[59] Petrochuk, M., & Zettlemoyer, L. (2018). Simplequestions nearly solved: A new upperbound and baseline approach. *arXiv preprint arXiv:1804.08798*.

[60] Wang, M., Smith, N. A., & Mitamura, T. (2007, June). What is the Jeopardy model? A quasi-synchronous grammar for QA. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)* (pp. 22-32).

[61] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).

[62] Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., ... & Petrov, S. (2019). Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, *7*, 453-466.

[63] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

[64] Tan, Y., Xu, H., Wu, Y., Zhang, Z., An, Y., Xiong, Y., & Wang, F. (2021). Research on knowledge driven intelligent question answering system for electric power customer service. *Procedia Computer Science*, *187*, 347-352.

[65] Junwen, L., Ziyan, Z., & Jiakai, H. (2017, February). The application of quantum communication technology used in electric power information & communication system confidential transmission. In *2017 19th International Conference on Advanced Communication Technology (ICACT)* (pp. 305-308). IEEE.

[66] Yin, L., Gao, Q., Zhao, L., Zhang, B., Wang, T., Li, S., & Liu, H. (2020). A review of machine learning for new generation smart dispatch in power systems. *Engineering Applications of Artificial Intelligence*, *88*, 103372.

[67] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.

[68] Santos, C. N. D., Xiang, B., & Zhou, B. (2015). Classifying relations by ranking with convolutional neural networks. *arXiv preprint arXiv:1504.06580*.

[69] Miwa, M., & Bansal, M. (2016). End-to-end relation extraction using lstms on sequences and tree structures. *arXiv preprint arXiv:1601.00770*.

[70] Zhu, H., Dong, L., Wei, F., Wang, W., Qin, B., & Liu, T. (2019). Learning to ask unanswerable questions for machine reading comprehension. *arXiv preprint arXiv:1906.06045*.

[71] Alberti, C., Andor, D., Pitler, E., Devlin, J., & Collins, M. (2019). Synthetic QA corpora generation with roundtrip consistency. *arXiv preprint arXiv:1906.05416*.

[72] Rajpurkar, P., Jia, R., & Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. *arXiv preprint arXiv:1806.03822*.

[73] Clark, C., Lee, K., Chang, M. W., Kwiatkowski, T., Collins, M., & Toutanova, K. (2019). BoolQ: Exploring the surprising difficulty of natural yes/no questions. *arXiv preprint arXiv:1905.10044*.

[74] Talmor, A., & Berant, J. (2018). Repartitioning of the complexwebquestions dataset. arXiv preprint arXiv:1807.09623.

[75] Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems (TOIS)*, *10*(1), 91-102.

[76] Taniuchi, D., & Maekawa, T. (2014, October). Robust Wi-Fi based indoor positioning with ensemble learning. In *2014 IEEE 10th International conference on wireless and mobile computing, networking and communications (WiMob)* (pp. 592-597). IEEE.

[77] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

[78] Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., ... & Socher, R. (2016, June). Ask me anything: Dynamic memory networks for natural language processing. In *International conference on machine learning* (pp. 1378-1387). PMLR.

[79] Cimiano, P., Unger, C., & McCrae, J. (2014). Ontology-based interpretation of natural language. Synthesis Lectures on Human Language Technologies, 7(2), 1-178.

[80] Green Jr, B. F., Wolf, A. K., Chomsky, C., & 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).

[81] Woods, W. A., & WA, W. (1977). Lunar rocks in natural English: Explorations in natural language question answering.

[82] Salunkhe, A. (2020). Evolution of techniques for question answering over knowledge base: A survey. International Journal of Computer Applications, 177(34), 9-14.

[83] Goldberg, Y. (2016). A primer on neural network models for natural language processing. Journal of Artificial Intelligence Research, 57, 345-420.

[84] Lende, S. P., & Raghuwanshi, M. M. (2016, February). Question answering system on education acts using NLP techniques. In 2016 world conference on futuristic trends in research and innovation for social welfare (Startup Conclave) (pp. 1-6). IEEE.

[85] Indurkhya, N., & Damerau, F. J. (2010). Handbook of natural language processing. Chapman and Hall/CRC.

[86] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.

[87] Reddy, S., Chen, D., & Manning, C. D. (2019). Coqa: A conversational question answering challenge. Transactions of the Association for Computational Linguistics, 7, 249-266.

[88] Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R., & Manning, C. D. (2018). HotpotQA: A dataset for diverse, explainable multi-hop question answering. arXiv preprint arXiv:1809.09600.