Question Answering Systems for Covid-19

Poonam Gupta¹, Ruchi Garg², Amandeep Kaur³

¹Assistant Professor, Computer Science and Engineering Department, MMEC, M.M.(Deemed to be) University, Mullana, Ambala, Haryana, India guptapoonam2311@gmail.com

²Assistant Professor, Electronics and Communication Engineering Department, MMEC, M.M. (Deemed to be) University, Mullana, Ambala, Haryana, India garg.330@gmail.com

³Associate Professor, Computer Science and Engineering Department, MMEC, M.M. (Deemed to be) University, Mullana, Ambala, Haryana, India amandeepkaur@mmumullana.org

Abstract. In the present scenario COVID-19 pandemic has ruined the entire world. This situation motivates the researchers to resolve the query raised by the people around the world in an efficient manner. However, less number of resources available in order to gain the information and knowledge about COVID-19 arises a need to evaluate the existing Question Answering (QA) systems on COVID-19. In this paper, we compare the various QA systems available in order to answer the questions raised by the people like doctors, medical researchers etc. related to corona virus. QA systems process the queries submitted in natural language to find the best relevant answer among all the candidate answers for the COVID-19 related questions. These systems utilize the text mining and information retrieval on COVID-19 literature. This paper describes the survey of QA systems- CovidQA, CAiRE (Center for Artificial Intelligence Research)-COVID system, CO-search semantic search engine, COVIDASK, RECORD (Research Engine for COVID Open Research Dataset) available for COVID-19. All these QA systems are also compared in terms of their significant parameters- like Precision at rank 1 (P@1), Recall at rank 3(R@3), Mean Reciprocal Rank(MRR), F1-Score, Exact Match(EM), Mean Average Precision, Score metric etc.; on which efficiency of these systems relies.

Keywords: COVID-19, Natural Language Processing, Question Answering, Text Mining, Information Retrieval
1. Introduction
Since the detection of SARS-CoV-2[1] or the Corona virus towards the end of Dec 2019 or the starting of 2020 the lives of people has drastically affected all over the world [2]. Therethere, doctors, health practitioners must keep up to date information about this coronavirus in order to save the lives of infected people to some extent. For this, Question Answering (QA) Systems has been a valuable source to find the accurate answers or the solutions to the questions during this COVID-19 pandemic. The effectiveness of the various question answering techniques is to find the most accurate answer out of multiple relevant answers. As the users tried to find out the information that is available online via search engines like Microsoft Bing, Yahoo, Google etc. the need for the automatic question answering based upon the domain COVID-19 becomes more urgent [3]. In order to get the exact information or an answer to a particular question has been a challenging task as large amount information about COVID-19 is available on Google and other search engines. In order to solve these issues various researchers uses the tools from Natural Language Processing, Machine Learning and Artificial Intelligence Using these tools QA systems try to get accurate answers for every different question instead of retrieving appropriate text documents[4]. In this paper we analyze the various question answering systems on COVID-19 based on natural language processing like CovidQA, CAiRE (Center for Artificial Intelligence Research)-COVID system, CO-search semantic search engine, COVIDASK, RECORD (Research Engine for COVID Open Research Dataset) etc. These question answering systems are tested using various covid-19 datasets available. The effectiveness of these systems is also compared using various metrics like F1-Score, Precision, Recall, Mean reciprocal Rank, opinions and then eliminate the words with highest frequency for each opinion, Exact Match (EM) etc.

2. Existing Question Answering Techniques for Covid-19

Raphael Tang et al.(2020)[4] presented CovidQA, the question answering dataset particularly designed to answer the questions related to COVID-19. CovidQA consists of 124 pairs of questions and documents. Using CovidQA, a number of techniques for transfer and unsupervised based question answering, number of transformer models was evaluated. Transformer based models comes out to be more effective when questions are in the form of natural language query rather than keyword based query. The input to CovidQA consists of (i) questions are put in the form of natural language query (ii) Complete text of ground truth article. Then the scoring of each sentence is done to get the accurate answer. To get an exact answer, pattern matching approach was used. Effectiveness of the answers are evaluated in terms of precision at rank one (P@1), mean reciprocal rank(MRR) and recall at rank three (R@3). Various models were compared like BM25, unsupervised neural techniques like BERT i.e. Bidirectional Encoder Representations from Transformers(Transformer)[6], SciBERT i.e. Bidirectional Encoder Representations from Transformers for scientific data [7], and BioBERT i.e. Bidirectional Encoder Representations from Transformers for Biomedical Text Mining)[8], BERT, BioBERT and T5 i.e. Text-to-Text transfer transformer[9] all are fine-tuned on MS MARCO(Machine Reading Comprehension Dataset)[10] and BioBERT (fine-tuned on SQuAD i.e. Stanford Question Answering Dataset)[11]. Among all these when the query is in the form of natural language, T5 achieves the highest overall effectiveness with precision at rank one value equals to 0.282[4], recall at rank three value equals to 0.404[4] and MRR value equals to 0.415[4]. When the query is keyword based then BERT fine tuned on MS MARCO performed well in terms of precision at rank one (P@1) with value equals to 0.234[4] and T5 performed well in terms of recall at rank three and MRR with values equals to 0.376[4] and 0.360[4] respectively.

Dan Su et al.(2020)[12] presented a CAiRE (Center for Artificial Intelligence Research)-COVID system consists of three major components (1) Document Retriever (2) Snippet Selector for relevant document and 3) Multi-Document Summarizer which is query-focused. Further, Document Retriever consists of two sub components Query Paraphrasing and Search Engine. The architecture of this
system is shown in ‘Figure 1’. First of all, to retrieve the most accurate document that contains the answer based upon the user query, document retriever component is used. The sub component of this module is used to rephrase the user query into some simple query, keep in view the meaning of the query remains the same. Then the sub component search engine is used to get the candidate documents for the query. Snippet Selector for relevant document is used to get the most accurate answer from the candidate answers. The Re-ranking score was calculated on the basis of Answer Confidence Score and Keyword-based Score. Answer Confidence Score is the prediction probability of the answer from the QA models. Keyword based Score is calculated by word matching between the query and the retrieved paragraphs for answers. Matching is done using POS tagging. Then the Re-ranking score is the summation of Answer Confidence Score and Keyword based Score. Multi-document Summarizer which is Query-focused is used to extract the abstractive summaries related to the queries of COVID-19. Matching is done using POS i.e., Part-of- Speech tagging.

![Figure 1. Architecture of CAiRE-COVID System](image)

Then the Re-ranking score is the summation of Answer Confidence Score and Keyword based Score. Query-focused Multi-document Summarizer is used to extract the abstractive summaries related to the queries of COVID-19.

Hillary Ngai et al. 2021[13], assessed the question-answering (QA) models on two labelled questions answers datasets CovidQA and CovidGQA using three suggested transformer-based question-answering systems BERT[6], ALBERT[14], and T5 model. The systems were pre-trained on various QA datasets like Stanford Question Answering Dataset (SQuAD) v1.1 [11], Stanford Natural Language Inference (SNLI) [15], Multi-Genre Natural Language Inference (MultiNLI) [16], Semantic Textual Similarity (STS) [17], Question-Answering Biomedical Dataset (BioASQ)[18] and evaluated on two labeled questions answers datasets — CovidGQA and CovidQA. The CovidGQA is a COVID-19 dataset created manually and encompasses 198 general question-text-answers related to COVID-19. The question-text has been taken out from medical websites, and medical subject-matter experts (SMEs) have provided answers, whereas, the CovidQA is a COVID-19 question-answering dataset developed manually from knowledge collected from the Kaggle CORD-19 dataset.
To extract each article’s relevant text to answer each question in the dataset, CovidQA dataset is combined with the CORD-19 dataset. The finally evaluated dataset contains 69 question-answer triplets. For the evaluation of QA systems, the evaluating metrics the macro-averaged F1 score and Exact Match (EM) of the answer extraction methods of QA systems on each dataset is used. BERT-large-uncased and ALBERT-base-uncased use whole word masking and are pre-trained on SQuAD v1.1, while T5-large was pre-trained on Colossal Clean Crawled Corpus (C4). The result obtained from the comparison of generated answer lengths on the CovidGQA dataset shows that BERT-large achieves the highest macro-averaged F1 score on both datasets, whereas, ALBERT-base unexpectedly outperforms BERT-large on EM, achieving the highest EM of the three models [13]. Although T5-large is having the most significant number of parameters (770M), both BERT-large and ALBERT-base outperform T5-large on all metrics on both datasets because T5-large was not pre-trained on a QA task. Based on the preliminary results obtained from BERT, ALBERT, and T5 model on two QA datasets and considering one of the significant limitations of transformers that they require a lot of labeled QA pairs to reach acceptable performance Hillary Ngai et al. 2021[13], proposed a hybrid QA system SBERT-BERT-QA and LDA-ALBERT-QA that combines few-shot learning with a transformer.

Estiva et al.[20] presented CO-search semantic search engine. It is a retrieval ranker based semantic search engine. Retrieval of documents is done using two keyword-based models and semantic model. Documents are embedded to create an index using pre-trained SBERT(Siamese-BERT)[21] used to embed the paragraphs. Further these embedding are combined with TF-IDF (Term Frequency-Inverse Document Frequency) and BM25. The linear combination of SBERT paragraph-level and TF-IDF document-level is used to retrieve the scores for each document. Then apply the reciprocal ranked fusion to combine this retrieved document with that obtained from BM25. Then the ranking is done with the help documents by the scores of retrieved documents, Question Answering module output and summarizer output. Various key metrics are used in evaluation like Precision, Bpref i.e. Binary Preference, MAP i.e. Mean Average Precision and nDCG i.e. After evaluation, CO-search semantic search engine ranked best among automated systems.

Lee et al.[22] presented COVIDASK question answering system. This system incorporates the biomedical text mining into existing question answering technique. It is used to provide the answers to questions in real time. The questions can be of the forms: interrogative questions (wh-type questions) like “where did covid new strain comes from?” or keyword based queries like “covid new strain come. Information Retrieval is also used for the evaluation purpose. Some factors are taken into account like Recency means for each question, up-to-date information or a document is to be used for the extraction of each answer. Another factor is how significant the resources are? i.e. answers are to be extracted from research papers that are published at venues of good reputation. The evidence document must be provided with each extracted answer. Latency must be minimum i.e. delay between inputting a question and receiving an answer must be minimum. In all the phrases contained in CORD-19[23] are pre-indexed and it is further used to build DENSPI(Dense Sparse Phrase Index)[24] model. It also used biomedical named entities described in PubMed for building BEST(Biomedical Entity Search tool)[25]. For a given query, COVIDASK returns different answers from both DENSPI and BEST. Firstly each vector of the phrase is supervised with an extractive dataset of Question Answering such as SQuAD[11]. Then all the candidate answer phrases are encoded into sparse and dense vectors. BEST makes inverted index and entity level search results are returned. BEST basically works well with keyword based queries. BEST is preferable over DENSPI. COVIDASK uses CORD-19 corpus related to coronavirus queries for the phrase indexing phase of DENSPI while BEST utilizes articles on PubMed. Two extractive datasets are used for training DENSPI. SQuAD[11] and natural questions dataset[2019]. Various types of questions like FAQ type questions coming from W.H.O, query log and KAGGLE’s CORD-19[19] challenge tasks are used to evaluate the performance of COVIDASK. Since COVIDASK implicitly performs information retrieval as it first returns the documents relevant to the answer for a given question. So
COVIDASK is also evaluated on IR tasks based i.e. TREC- COVID [27][28] Challenge. The performance is superior on interrogative questions, when DENSPI + SPARC is trained on SQuAD. DENSPI which is trained on natural questions outperforms well upon keyword-based questions. The results on TREC-COVID shows that COVIDASK is not very effective when compared with the other systems that are fully dedicated to information retrieval rather than question answering. COVIDASK efficiently handles the natural language questions both in keyword and interrogative forms.

Lu et.al [29] proposed a pipeline to describe what the CORD-19 articles revealed by automatically return the answers for the questions related to COVID-19. Then the aggregation of the answers is done. This pipeline consists of various modules: (i) Context Retrieval Module, it is used for finding out the only the relevant passages from the entire documents. (ii) Question-Answering Model, given a question and passage chosen by context retrieval module BERT (Bidirectional encoder representations from transformers) [6] is used for the prediction of location and the length of the answer in each passage as shown in ‘Figure 2’.

Muffo et al.[30] presented RECORD i.e. Research Engine for COVID Open Research Dataset. It is the tool to provide the answers to the questions based on COVID-19. It uses CORD-19 dataset, focused only on the articles about COVID-19 and SARS-CoV-2. Three steps are involved in this system: (i) Preprocessing Step used to split the entire text into chunks that are related to query posed by the user. Entire text is broken down into paragraphs in such a way so that semantic relationship between the query and the extracted paragraphs are not broken. (ii) Embedding Step used to embed all the chunks in the body text. Basically, the query and the related chunks both are embedded using Sentence-BERT model[21]. Cosine similarity is calculated between sentence embedding based on query and sentence embedding based on the chunks. It is used extract the most semantically related text documents to the query. (iii) Question Answering Step used to extract the answer. Question answering model based on the architecture of BERTLARGE is used. This step used to provide the best answer among all the chunks along with the score. For each chunk, a score is calculated then among all the chunks, a chunk
that has the highest score is considered as a best answer among all the related answers. Score is used as a discriminating factor among all the related chunks for a given query. RECORD also provides the title, author, publishing journal, paper citations, Scimago Journal Score etc. The summarization table of different question answering systems on COVID-19 is given in Table 1.

| S.No. | Question Answering System on Covid-19 | Metrics Used for Evaluation | Effectiveness |
|------|--------------------------------------|-----------------------------|---------------|
| 1.   | CovidQA by Raphel Tang et al.[4]     | Precision at rank 1 (P@1), Recall at rank 3(R@3), Mean Reciprocal Rank(MRR) | T5 is highly effective on all natural language questions while BERT is effective in terms of P@1 when the queries are in the form of keyworda |
| 2.   | CAiRE (Center for Artificial Intelligence Research)-COVID system by Dan Su et al.[12] | Precision at rank 1 (P@1), Recall at rank 3(R@3), Mean Reciprocal Rank(MRR) | CAiRE outperforms well. |
| 3.   | Transformer- based question-answering systems BERT[6], ALBERT[14], and T5 model by Hillary Nigai et al.[13] | F1-Score and Exact Match(EM) | BERT-large achieves the highest F1-Score on both datasets on CovidQA and CovidGQA. ALBERT-base achieves highest EM value. |
| 4.   | CO-search semantic search engine by Estiva et al.[20] | Normalized discounted cumulative gain(nDCG), precision, mean average precision (MAP) and binary preference(Bpref) | CO-Search, COVID- 19 scientific search engine ranked best among all automated systems. |
| 5.   | COVIDASK Question Answering System by Lee et al.[22] | in-sentence EM (EMsent) and top-k EMsent (EMsent@k) | COVIDASK when trained on SQuAD and Natural Questions outperforms well. |
| 6.   | Question Answering System based on context Retriever, Question Answering Model and Opinion Aggregator by Lu et al.[29] | Opinions and the words with the highest frequency for each opinion | Proposed method provides useful insights into COVID-19 related questions along with opinions and captures the recent trends in research about COVID-19. |
| 7.   | RECORD Question Answering System by Muffo et al.[30] | Score metric in terms of precision and consistency with input question | RECORD outperforms well in providing consistent and precise answer. |

3. Conclusion
In the recent pandemic situation QA systems can help each of its users to acquire relevant information effectively. However, most of the users like to submit their queries in their natural languages for
processing to these systems. Therefore, QA systems are designed and modeled to process queries by using natural language processors. The performance of these QA systems can be analyzed by using various metrics like Precision at rank 1 (P@1), Recall at rank 3 (R@3), Mean Reciprocal Rank (MRR), F1-Score, Exact Match (EM), Mean Average Precision, Score Metric as discussed. On the basis of these metrics we observe that some QA systems are efficient to process queries on the basis of keywords matching like BERT while others on the basis of their semantic analysis like T25. At the same time SBERT with combination TF-IDF and BM25 shows a remarkable efficiency to process queries by using keywords matching as well as semantic analysis. It implies that QA systems should process the queries by using both i.e. keyword matching and semantic analysis. Conclusively it is observed that CO-search semantic search engine based on SBERT with combination of TF-IDF and BM25 is more preferred QA system in recent times.

References

[1] Wu F, Zhao S, Yu B, et al. “A new coronavirus associated with human respiratory disease in China”, Nature 579, pp. 265–269, 2020.

[2] Wang Lucy Lu and Lo Kyle, “Text mining approaches for dealing with the rapidly expanding literature on COVID-19”, Briefings in Bioinformatics, Volume 22, Issue 2, https://doi.org/10.1093/bib/bbaa296, pp 781–799, March 2021.

[3] Gupta Poonam, Gupta Vishal, “Hybrid Approach for Punjabi Question Answering System”, In: Thampi S., Gelbukh A., Mukhopadhyay J. (eds) Advances in Signal Processing and Intelligent Recognition Systems, Advances in Intelligent Systems and Computing, vol 264, Springer, Cham. https://doi.org/10.1007/978-3-319-04960-1_12, 2014, pp 133-149.

[4] Gupta Poonam and Gupta Vishal, “A Survey of Existing Question Answering Techniques for Indian Languages”, Journal of Emerging Technologies in Web Excellence, Vol 6, 2014, pp. 165-168.

[5] Tang Raphael, Nogueira Rodrigo, Zhang Edwin, Gupta Nikhil, Cam Phuong, Cho Kyunghyun, and Lin Jimmy, Rapidly bootstrapping a question answering dataset for COVID-19. CoRR, abs/2004.11339, 2020, pp 1-8.

[6] Devlin Jacob, Chang Wei-Ming, Lee Kenton, and Toutanova Kristina, “BERT: Pre-training of deep bidirectional transformers for language understanding”, In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Minneapolis, Minnesota, 2019, pp. 4171–4186.

[7] Beltagy I, Lo Kyle, and Cohan Arman, "SciBERT: A pretrained language model for scientific text”, 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), 2019, pp. 3615-3620.

[8] Lee Jinhyuk, Wonjin Yoon Wonjin, Kim Sungdong, Kim Donghyeon, Kim Sunkyu, Ho So Ho Chan, and Kang Jaewoo,"BioBERT: a pretrained biomedical language representation model for biomedical text mining".Bioinformatics, Vol. 36, Issue 4, 2020, pp. 1234-1240.

[9] Raffel Colin, Shazeer Noam, Roberts Adam, Lee Katherine, Narang Sharan, Matena Michael, Zhou Yanqi, Li Wei, and Liu J. Peter, “Exploring the limits of transfer learning with a unified text-to-text transformer”.Journal of Machine Learning Research, In arXiv:1910.10683, 2019,
Bajaj Payal, Campos Daniel, Craswell Nick, Deng Li, Gao Jianfeng, Liu Xiaodong, Majumder Rangan, McNamara Andrew, Mitra Bhaskar, Nguyen Tri, Rosenberg Mir, Song Xia, Stoica Alina, Tiwary Saurabh, and Wang Tong, “MS MARCO: A human generated MAchine Reading COMprehension dataset”, 30th Conference on Neural Information Processing System (NIPS), arXiv:1611.09268, 2016, pp. 1-11.

Rajpurkar Pranav, Zhang Jian, Lopyrev Konstantin, and Percy Liang Percy, “SQuAD: 100,000+ questions for machine comprehension of text”, In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for computational Linguistics, 2016, pp. 2383-2392.

Dan Su, Yan Xu, Tiezheng Yu, Farhad Bin Siddique, Elham J. Barezi, and Pascale Fung, ”Caire-covid: A question answering and multi-document summarization system for COVID-19 research.”, CoRR, abs/2005.03975, 2020, pp. 1-11.

Ngai Hillary, Park Yoona, Chen John, Parsapoor Mahboobeh, “Transformer-Based Models For Question Answering On Covid19”, arXiv:2101.1143, 2021, pp. 1-7.

Bowman R. Samuel, Nangia Nikita, and Bowman Samuel, “A broad-coverage challenge corpus for sentence understanding through inference”, In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, Association for Computational Linguistics, 2018, pp. 1112-1122.

Guo Xiao, Mirzaalian Hengameh, SabirEkraam, Jaiswal Ayush, Wael Abd-Almageed, “CORD19STS: COVID-19 Semantic Textual Similarity Dataset”, 2020 pp 1-8.

Tsatsaronis, G., Balikas, G., Malakasiotis, P. et al., “An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition”, BMC Bioinformatics 16:138, https://doi.org/10.1186/s12859-015-0564-6, 2015 pp. 1-28.

COVID-19 Open Research Dataset Challenge: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge.

E steva, A., Kale, A., Paulus, R. et al. “COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization”. npj Digit. Med. 4, 68, https://doi.org/10.1038/s41746-021-00437-0, 2021 pp. 1-10.

Nils Reimers and Iryna Gurevych. “Sentence-Bert: Sentence embeddings using Siamese Bert networks”, 2019, pp. 1-11.
[22] Lee Jinhyuk, Yi S. Sean, JeongMinbyul, Sung Mujeen, Yoon Wonjin, Choi Yonghwa, Ko Miyoung, and Kang Jaewoo, “Answering Questions on COVID-19 in Real-Time”, Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP, Association for Computational Linguistics, doi: 10.18653/v1/2020.nlpcovid19-2.1, 2020 pp.1-10.

[23] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jianguang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. “Cord-19: The covid-19 open research dataset”, arXiv preprint arXiv:2004.10706, 2020, pp. 1-12.

[24] SeoMinjoon, LeeJinhyuk, Kwiatkowski Tom, Parikh Ankur, Farhadi Ali, HajishirziHannaneh, “Real-Time Open-Domain Question Answering with Dense-Sparse Phrase Index”, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp: 4430–4441, Association for Computational Linguistics, doi: 10.18653/v1/P19-1436 ,2019, pp. 4430-4441.

[25] LeeSunwon, KimDonghyeon, LeeKyunbum, ChoiJaehoon, KimSeongsoo, Jeon Minji, LimSangrak, ChoiDonghee, Kim Sunkyu, Tan Aik-Choon, et al.,” Best: next-generation biomedical entity search tool for knowledge discovery from biomedical literature”, PloS one, 11(10), 2016, pp. 1-16.

[26] Kwiatkowski Tom, PalomakiJennimaria, Redfield Olivia, Collins Michael, Parikh Ankur, Alberti Chris, Epstein Danielle, PolosukhinIllia, Devlin Jacob, Lee Kenton, et al.” Natural questions: a benchmark for question answering research”, Transactions of the Association for Computational Linguistics, 7, 2019, pp.453–466.

[27] Roberts Kirk, AlamTasmeer, Bedrick Steven, Fushman-Demner Dina, Lo Kyle, Soboroff Ian, Voorhees Ellen, Wang Lu Lucyand Hersh R William, “Trec-covid: Rationale and structure of an information retrieval shared task for covid-19”, Journal of the American Medical Informatics Association,Vol. 27, Issue 9,2020, pp. 1431-1436.

[28] Voorhees Ellen, AlamTasmeer, BedrickSteven, Fushman-Demner Dina, Hersh R William, Lo Kyle, Roberts Kirk, Soboroff Ian, and Wang Lu Lucy, ”Trec-covid: Constructing a pandemic information retrieval test collection”,arXiv preprint arXiv:2005.04474,pp.1-10, 2020.

[29] Lu Z, Wang JX and Li X, Revealing Opinions for COVID-19 Questions Using a Context Retriever, Opinion Aggregator, and Question-Answering Model: Model Development StudyJ Med Internet Res 2021;23(3):e22860, doi:10.2196/22860.

[30] Muffo Matteo, Cocco Aldo, Messina Mattia and Bertino Enrico,” RECORD: a Question Answering tool for COVID-19”, EMNLP 2020 Workshop NLP-COVID Submission, 2020.