Feasibility Study of a BERT-based Question Answering Chatbot for Information Retrieval from Construction Specifications

J. Kim, S. Chung, S. Moon, and S. Chi
Department of Civil & Environmental Engineering, Seoul National University, Seoul, Korea
(janykjy@snu.ac.kr, hwani751@snu.ac.kr, blank54@snu.ac.kr, shchi@snu.ac.kr)

2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) | 978-1-6654-8687-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEM55944.2022.9989625

Abstract - **Checking construction specification in every construction phase is critical to ensure proper construction quality and to avoid contractual problems. However, manual review is inefficient, expensive, and error-prone. There have been efforts to automatically review specifications, but these studies are limited in their practical applicability. As a solution, the use of retrieval-based user interface (as known as a chatbot) can extract specific information from construction specifications as a user wants. For the development of an information retrieval chatbot for construction specifications, this paper tested the application feasibility of a question answering methodology using Bidirectional Encoder Representations from Transformers (BERT). By taking advantages of the pre-trained BERT, user-wanted information was successfully extracted from construction specifications. With this approach, variety of questions can be responded flexibly without time-consuming manual tasks such as labeling.**

Keywords - **Bidirectional Encoder Representations from Transformers (BERT), Chatbot, Construction Specification, Information Retrieval (IR), Question Answering (QA)**

I. INTRODUCTION

Construction specifications are the documents written by clients, describing how construction project should be carried out. Specifications usually include detailed description about dimensions, materials, construction, quantities, and others of projected work, which cannot be expressed through drawings but should be written in words. Therefore, a considerable portion of the document is composed of a textual form. In every construction phase, it is crucial to review construction specifications because these documents are legally binding contracts [1].

Extracting information from construction specifications in an objective manner can help project participants to ensure proper construction quality and avoid contractual problems. For example, in the early phase of the project, it is necessary to examine construction specifications in order to check if all requirements are appropriate for the site environment [2]. When bidding on construction projects, to properly manage project risks, understanding of such documents is particularly important [1]. In quality compliance checking, regulations play critical role in assuring construction quality [3]. Also in execution phase, accurate recognition of constraints and standards can prevent delays or rework. Thus, factual information contained in construction specifications is required at different stages of the project, and reviewing such documents is fundamental for the successful delivery of construction projects.

Construction specifications are usually composed in a semi-structured form in which they are divided by sections. However, due to the amount of information that tends to exceed over thousands of pages in general, it requires significant human effort to review these documents, which is time-consuming, costly, and error-prone. This makes it necessary to develop a model to automate the process of reviewing the documents. However, variance in section configuration of different specifications makes it difficult to use these structures for developing such a model. Therefore, when developing a model that searches for specific information from the specifications, it is recommended to find them by using content of the information rather than the structure of the documents.

Various efforts have been made to extract specific information from different construction documents in an automated way ([1], [7], [8], [9], [10]). But these studies are limited in that they require a lot of manual input when developing the models, such as labeling or rule development. Also, there is a limit in applicability that they could only be applied to specific types of documents. To extract new type of information, new rules should be developed, or models should be trained on additional labeled dataset.

Therefore, a retrieval-based user interface (as known as a chatbot) that searches based on the content is necessary for providing information from construction specifications. Recently, chatbots have been gaining attention, and many companies have been launching chatbot services for various purposes (e.g., personal security, work support, and customer service). Retrieval-based chatbots can quickly find and provide users with needed information. Because a retrieval-based chatbot is not rule-based, it is flexible in responding to new types of questions. However, there were no existing cases of such chatbots used for retrieving information from construction specifications. Cho and Lee (2019) have adopted a chatbot for managing construction daily reports, which collects information through conversations and generates and shares daily reports for project participants [4]. However, the proposed chatbot is rule-based, leading users to pre-defined patterns of conversation. Therefore, only the pre-defined type of information can be enrolled and extracted.

When the retrieval-based chatbot is properly used, it can flexibly respond to unseen questions. In line with the limitations of the previous studies and the characteristics of construction specifications, needed information from construction specifications can be well retrieved using a methodology such as Question Answering (QA). QA, one of the natural language processing (NLP) tasks, aims to retrieve small fragments of text which contain the answer to a question from a given context [5]. Bidirectional Encoder Representations from Transformers (BERT), a state-of-the-art method for QA, is a Transformer-based language model pretrained on a large number of unlabeled text data. Weights of BERT are then adjusted with additional training data for NLP tasks, which is called "fine-tuning". BERT-based approaches have demonstrated outstanding performance in range of NLP tasks, including QA [6].

Therefore, this research aims to demonstrate the feasibility of BERT-based QA method towards the development of an information-retrieving chatbot for construction specification. The two-staged QA method based on pre-trained BERT was used to search for information from specifications. In the first stage, the BERT-based model was fine-tuned to retrieve relevant paragraphs (i.e., paragraphs that contain the answer to a given question), and in the second stage, BERT fine-tuned on general domain dataset for QA tasks was adopted to extract the answer from the relevant paragraph. A feasibility study was conducted with the case of *Standard Specifications of Department of Transportation of New York*. For the feasibility study, a dataset of paragraph-question-answer triplets was constructed, and the performance of the BERT model was evaluated.

The rest of the document is organized as follows: Section 2 presents the literature review on information extraction and retrieval in construction documents and QA. In section 3, the methodology for QA and setups for feasibility study are explained, followed by the results, and discussion in Section 4. Section 5 concludes the study, presents the limitations, and suggests the future study.

II. LITERATURE REVIEW

A. *Information Extraction and Retrieval in Construction Documents*

There have been a lot of effort to extract important information from construction documents using NLP and text mining methodologies in academia. Zhang and El-Gohary [7] proposed a domain-specific, semantic, rule-based NLP approach for automated information extraction from construction regulatory documents. In their study, pattern-matching rules to extract information were manually built. Later, Zhang and El-Gohary [8] proposed a fully automated compliance checking (ACC) system. Their ACC system applied semantic, logic-based representations to extract regulatory information and design information from regulatory documents and from BIM, respectively. Akanbi and Zhang [9] endeavored to extract design information from construction specification to automate the cost estimation process, by using semantic modeling and NLP techniques. Their study also adopted a pattern extraction method, which is rule-based.

Previous studies were based on manually-built rules to analyze construction documents. These studies identified the types of information to be extracted, found patterns, and defined rules that extracts information. Although the previous studies derived good performance in information extraction, their methods take considerable amount of time and cost to be implemented. In addition, proposed pattern-matching rules can only be applied to extracting pre-defined types of information from specific types of documents. If a target document or information changes, new rules need to be built.

There exist studies adopting machine learning and deep learning approaches. The study by Kim and Chi [10] proposed a knowledge management system that retrieves accident cases and analyzes tacit knowledge from the construction accident case database. To retrieve similar accident cases, Okapi BM25, one of the variations of term frequency and inverse document frequency (TF-IDF), was used, and the thesaurus-based weighting scheme was applied. Accident cases were ranked by their similarity with weighted query, and knowledge was extracted by the rule-based method and the conditional random field (CRF) approach. Moon et al. [1] proposed a named entity recognition (NER) approach based on bidirectional long short-term memory (Bi-LSTM) architecture, which was applied to automatic review of construction specification at the bidding stage. Five information types were defined, which referred to five named entities: organization, action, element, standard, and reference. In their study, manual labeling work was inevitable for the training of the NER model.

These studies adopted machine learning and deep learning for automatic information retrieval and extraction, but their methods still require excessive manual effort in constructing thesaurus or in developing labeled training data. Additional effort is also required when new concepts or information arise.

B. *Question Answering*

Question answering (QA) is the NLP task to retrieve the answer to a question from the given passage. It is an information retrieval task, and recently have gained attention with the use of deep neural networks [11].

In particular, many QA studies have been conducted through BERT model. Karpukhin et al. [12] conducted the study in the open-domain QA, which is the task to find answers from a large set of documents with a variety of topics. The study showed that retrieval can be implemented through dense representations, where embeddings are learned from the set of questions and passages by a dual encoder framework. With the dual encoder composed of pre-trained BERT, retrieval was conducted by calculating the similarity between vectors of question and passage. Kim et al. [13] proposed a QA method to retrieve infrastructure damage reports and to provide users with the information from the retrieved reports. 435 question-answer pairs were used to fine-tune a BERT model to acquire QA capability in context of infrastructure damage. However, up to the authors' best knowledge, the QA method has not been applied to finding necessary contents from construction specifications. Therefore, this research aims to test whether QA is applicable to searching for information from construction specifications.

III. METHODOLOGY

This study tests the feasibility of QA methodology for information retrieval and extraction from construction specifications. The methodology adopted in this study is shown in Figure I.

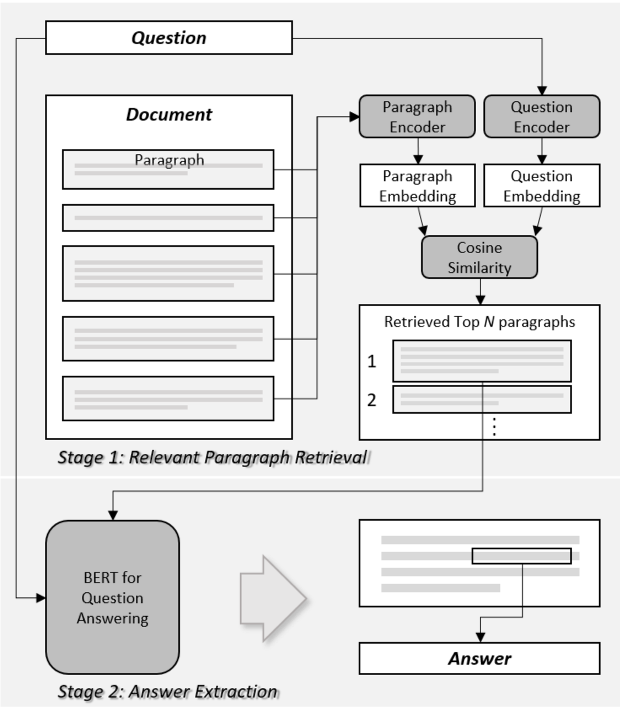

Figure I. Two-staged methodology of QA

A. Stage 1: Relevant Paragraph Retrieval

The first stage is to retrieve the relevant paragraphs which are likely to contain the answer to the given question. Because the answer can be found only when the correct context is provided, this stage is critical to the performance of QA at the second stage.

The process of paragraph retrieval is as follows. Two encoders are needed for paragraph embedding and question embedding, respectively. Paragraph embedding expresses an entire paragraph as a vector, and question embedding expresses a question as a vector. When a question is given, every paragraph in a target document and the question are embedded into vectors by each encoder. Similarity between two vectors is calculated by cosine similarity, the dot product of two embedding vectors. Finally, top N paragraphs with the highest similarity scores are retrieved.

Each encoder (i.e., paragraph encoder and question encoder) is trained by adjusting embedding vectors to increase the scores of relevant questions and paragraphs. Therefore, question-paragraph pairs labeled as 'relevant' are needed for the training. As a result, high similarity score is returned for relevant paragraphs and questions, and low similarity score is returned for irrelevant paragraphs and questions.

In this study, a BERT-based encoder was used for paragraph and question embeddings. Encoders from the Hugging Face Transformers library were imported and then fine-tuned on Standford Question Answering Dataset (SQuAD). SQuAD is a reading comprehension dataset, consisting of triplets of the paragraph, the question, and the answer.

B. Stage 2: Answer Extraction

The following step is to search for the answer from the candidate paragraph resulted from the stage 1. In the second stage, the answer is directly derived from the paragraph. Therefore, it is important to find the location of the answer, where it begins and ends.

In this study, another BERT model was used. To minimize the manual effort, the BERT model already fine-tuned on SQuAD was imported from the Hugging Face. This BERT model is fed by two inputs; the question and the paragraph. Then, the model extracts the excerpts from the paragraph as the answer.

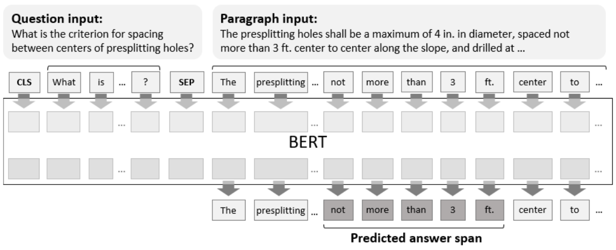

Figure II. Details of answer extraction

Details are as follows. Two inputs (i.e., the question and the paragraph) are fed into the fine-tuned BERT. Tokenization of the inputs is conducted, and special tokens [CLS] and [SEP] are used to indicate the beginning of a sentence and the separation between two inputs, respectively. After the tokenization, at the last layer of the BERT, a sequence of two-dimensional vectors is generated for every token in the paragraph. The first element of a two-dimensional vector represents the probability that the corresponding token is the start of the answer, and the second element represents the probability that the corresponding token is the end of the answer. By referring to the output vectors, the model finds the start and end tokens of the answer. As a result (Figure II), the model can finally extract the answer tokens spanning from the start token to the end token.

C. Scope of the Feasibility Study

As the first step of testing the feasibility of applying the two-staged QA methodology to extract required information from construction specifications, this study tested the performance of answer extraction (stage 2). A fine-tuned BERT was used, context and question data were fed to the model, and the predicted answer was extracted as output. Evaluation metrics of exact match, precision, recall, and F1 score were used to test the feasibility of the BERT-based model.

IV. EXPERIMENT

A. Dataset Construction

Dataset to test the QA methodology was created from the *Standard Specifications of road Department of Transportation of New York*. The dataset contains 54 triplets of context, question, and answer. Random paragraphs in the specification were selected as 'context' data. At least one set of question and answer pairs for each paragraph was formed.

The authors adopted the types of information that should be retrieved in construction specifications (Table I) from the study by Moon et al. [1], and defined another information type 'Definition and explanation of terms'. The following is question type to extract the defined information types from construction specification:

1. "Who is responsible?"
2. "What should be extracted by When?"
3. "How should be done by when?"
4. "Which quality standards should be followed?"
5. How can the term be defined?

TABLE I
INFORMATION TYPES IN CONSTRUCTION SPECIFICATION

	Information types defined	Numbers of questions generated
1	People and organizations in charge	13
2	Quality standards and criteria	22
3	Activities required	8
4	Relevant references	2
5	Definition and explanation of terms	9
Total		54

Finally, aforementioned dataset consisting of 54 triplets of context, question, and ground truth answer was generated belonging to one of the information types defined. The numbers of generated questions are shown in Table I, and examples of the questions and the answers are shown in Table II. The created dataset was used to test the performance of the BERT-based model if the defined types of information can be well retrieved through the QA method.

TABLE II
EXAMPLES OF CONSTRUCTED DATA

Context & ***Answer***[a,b]		Question
Curing Covers. Use of curing covers is subject to the approval of ***the Engineer***[a]. Use quilted covers, plastic coated fiber blankets, or polyethylene curing covers. Do not use ***covers with tears or holes***[b]. Cover all exposed surfaces and extend the covers a minimum of 12 inches beyond the pavement edges or beyond the forms, when used. Overlap successive covers 12 inches, minimum. Secure the covers to keep them in contact with the entire surface and maintain the overlap. Wet the entire surface of quilted covers and maintain them in a wetted condition throughout the curing period.	**a**	Whose approval is required for the use of curing covers?
	b	What covers are restricted from being used?

B. Results and Discussion

The results showed that 36 answers out of 54 exactly matched with the ground truth answers, achieving 0.6667 of accuracy. Considering the fact that the only exact matches were regarded correct, this result is quite promising. In terms of precision, recall, and F1 score, the model achieved 0.9765, 0.7847, 0.8702, respectively.

TABLE III
EXAMPLES OF THE INCORRECT ANSWER PREDICTED BY THE MODEL

	Question	Ground truth answer containing ***Predicted answer***
1	Where should nonfriable asbestos be disposed?	permitted ***C&D waste management facility***
2	What is the unacceptable difference between the top of pile driven into the ground at full length and its plan location?	more than ***4 inches***
3	In option A, how many days does it take for Geotechnical Engineering Bureau to take action after receiving the request and sample?	at least ***fourteen*** days
4	How should the water supply line be connected with a valve immediately downstream of the flow meter?	a ***90° T*** connection

The authors examined wrong answer cases. Examples of wrongly predicted answers are shown in Table III. Most of the incorrect cases included at least a fraction of the answer. Especially, high precision of 0.9765 show that the model well locates the main keywords of the correct answer, without any error case where the model predicted a completely different answer. So, it could be concluded that the model can find the location of information very accurately that questioner wants to know. This is a meaningful result in that even a model that is not pre-trained on construction corpus can roughly locate the required content in construction specification documents.

It should be noted that some words omitted from the answer may vary the meaning of the answer completely (i.e., first three rows of Table III). This shows that the model should be improved to extract answers with wider span. Nevertheless, in the second and third error cases, even though some information was omitted, it was found not to be critical when inferring the meaning. For example, when a project participant asked a question about certain criteria, the most notable part will be the specific number (i.e., 4 inches, fourteen). In other words, most of the participants at the construction site know how the construction is done, so they will be able to determine whether the preceding word the model omitted is "less than" or "more than".

As shown above, the BERT-based QA model successfully extracted the answer from a given paragraph. It is worth noting that about 30% of the predicted answers did not exactly match with the correct answer, but even in these cases, they included at least some parts of the correct answer and the extraction of key information was successful. For these reasons, the results of this study show that BERT-based QA can be applied to extracting information from the construction specification.

V. CONCLUSION

This study validated feasibility of the BERT-based QA method for developing a retrieval-chatbot for construction specifications. A pilot study was conducted with the data built with construction specification document of New York. The experiment identified whether the model finds corresponding answer to the question from the paragraph. An exact match accuracy was measured as 0.6667, which is not very low despite the use of strict evaluation method. Also, considering the result of precision, recall and F1 score, the results are promising given that model is competent in finding fraction of the ground truth answer.

The study discovered meaningful result that the model pre-trained only on general language corpus could be utilized in domain specific search. Also, to the best of our knowledge, there was no previous attempt in which BERT was adopted to searching for information from construction specifications.

Limitation exists that the dataset was created through only one construction specification. To prove the model's usefulness, it should be tested on other datasets as well. However, the researcher expects that it will show good performance in other specifications as well because the model for general QA have worked well in specification of New York.

The study conducted a feasibility study only to the answer extraction stage. Therefore, in future work, study for paragraph retrieval (stage 1) should be carried out. In the extraction stage, it should be also tested on other construction specifications as well. As discussed above, the answer extracting model can be trained to spit out more comprehensive answers.

ACKNOWLEDGMENT

This research was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0008475, Development Program for Smart Digital Engineering Specialist)

REFERENCES

[1] S. Moon, G. Lee, S. Chi, H. Oh, "Automated construction specification review with named entity recognition using natural language processing," *Journal of Construction Engineering and Management*, vol. 147, no. 1, Jan. 2021.

[2] S. Moon, G. Lee, S. Chi, "Semantic text-pairing for relevant provision identification in construction specification reviews," *Automation in construction*, vol. 126. Aug. 2021.

[3] B.T. Zhong, L.Y. Ding, H.B. Luo, Y. Zhou, Y.Z. Hu, H.M. Hu, "Ontology-based semantic modeling of regulation constraint for automated construction quality," *Automation in Construction*, vol. 28, pp 58-70, Dec. 2012.

[4] J. Cho, G. Lee, "A chatbot system for construction daily report information management," in *Proc. 36th ISARC, Banff, Canada*, pp. 429-437.

[5] E. M. Voorhees, "The TREC-8 question answering track report", *TREC*, vol. 99, pp. 77-82, 1999.

[6] O. Kovaleva, A. Romanov, A. Rogers, A. Rumshisky, "Revealing the dark secrets of BERT," unpublished

[7] J. Zhang, N. M. El-Gohary, "Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking," *Journal of Computing in Civil Engineering*, vol. 30, no. 2, Mar. 2016.

[8] J. Zhang, N. M. El-Gohary, "Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking," *Automation in Construction*, vol. 73, pp. 45-57, Jan. 2017.

[9] T. Akanbi, J. Zhang, "Design information extraction from construction specifications to support cost estimation," *Automation in Construction*, vol. 131, Nov. 2021.

[10] T. Kim, S. Chi, "Accident case retrieval and analyses: Using natural language processing in the construction industry," *Journal of Construction Engineering and Management*, vol. 145, no. 3, Mar. 2019.

[11] Z. Abbasiantaeb, S. Momtazi, "Text-based question answering from information retrieval and deep neural network perspectives: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 6, May, 2021.

[12] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, W. Yih, "Dense passage retrieval for open-domain question answering," unpublished

[13] Y. Kim, "Question answering method for infrastructure damage information retrieval from textual data using bidirectional encoder representations from transformers," *Automation in Construction*, vol. 134, Feb. 2022.