sustainability

MDPI

Article

Deep Learning-Based Knowledge Graph Generation for COVID-19

Taejin Kim, Yeoil Yun and Namgyu Kim *

Graduate School of Business IT, Kookmin University, Seoul 02707, Korea; jeung722@kookmin.ac.kr (T.K.); yunyi94@kookmin.ac.kr (Y.Y.)
* Correspondence: ngkim@kookmin.ac.kr; Tel.: +82-2-910-5425; Fax: +82-2-910-4017

Abstract: Many attempts have been made to construct new domain-specific knowledge graphs using the existing knowledge base of various domains. However, traditional "dictionary-based" or "supervised" knowledge graph building methods rely on predefined human-annotated resources of entities and their relationships. The cost of creating human-annotated resources is high in terms of both time and effort. This means that relying on human-annotated resources will not allow rapid adaptability in describing new knowledge when domain-specific information is added or updated very frequently, such as with the recent coronavirus disease-19 (COVID-19) pandemic situation. Therefore, in this study, we propose an Open Information Extraction (OpenIE) system based on unsupervised learning without a pre-built dataset. The proposed method obtains knowledge from a vast amount of text documents about COVID-19 rather than a general knowledge base and add this to the existing knowledge graph. First, we constructed a COVID-19 entity dictionary, and then we scraped a large text dataset related to COVID-19. Next, we constructed a COVID-19 perspective language model by fine-tuning the bidirectional encoder representations from transformer (BERT) pre-trained language model. Finally, we defined a new COVID-19-specific knowledge base by extracting connecting words between COVID-19 entities using the BERT self-attention weight from COVID-19 sentences. Experimental results demonstrated that the proposed Co-BERT model outperforms the original BERT in terms of mask prediction accuracy and metric for evaluation of translation with explicit ordering (METEOR) score.

Keywords: deep learning; knowledge graph; text analytics; pre-trained language model; BERT

Citation: Kim, T.; Yun, Y.; Kim, N. Deep Learning-Based Knowledge Graph Generation for COVID-19. *Sustainability* **2021**, *13*, 2276. https://doi.org/10.3390/su13042276

Academic Editor: Piera Centobelli
Received: 27 November 2020
Accepted: 10 February 2021
Published: 19 February 2021

1. Introduction

Due to the continuous spread of COVID-19, the demand for information from various coronavirus-related domains has increased significantly. Such information is usually generated through various media, papers, and patents, and the amount of new information is increasing quickly. As information related to the coronavirus is accumulated continuously, several methods have been considered to manage this information effectively. These methods typically extract meaningful information from huge datasets and refine the information to applicable knowledge. Specifically, these methods have been used to construct new knowledge bases by grasping the intrinsic relationships between entities from a raw dataset and representing these relationships as refined knowledge or by applying an existing knowledge base to represent new relationships. In addition, many attempts have been made to construct knowledge graphs (KGs) that represent knowledge as a directional network.

The original concept of a knowledge graph was introduced in 1972 [1], and, in 2012, Google applied this concept to improve the performance of its search engine by constructing a new knowledge graph to create a knowledge graph-based semantic database [2]. A

semantic database is a kind of knowledge base that declares relationships between various entities included in the database as a triple structure and defines entity concepts by these relationships. Semantic databases are used in various fields, such as ontology and the Semantic Web. However, differing from a traditionally structured semantic database, a knowledge graph can easily join new entities or expand new relationships because it is, essentially, a network representation of a knowledge base. For example, new knowledge items extracted from external sources can be created and existing items can be modified, updated, or removed.

This implies that a KG is easy to update and has excellent expandability. However, traditional "dictionary-based" or "supervised" knowledge graph construction methods will not be able to adapt to new vocabulary in new problem domains. This limitation is emphasized when constructing a domain-focused knowledge graph that cannot be explained with a pre-built dictionary or dataset, and it is nearly impossible to define knowledge in cases where information can be added or updated frequently, such as in the case of the recent COVID-19 pandemic. Figure 1 shows how this limitation affects the construction of a COVID-19-specific knowledge graph based on a traditional knowledge graph. Therefore, OpenIE knowledge graph construction techniques prove to be an effective solution to the problem of knowledge graph generation in the face of evolving knowledge, as they do not require expensive efforts to annotate new datasets. In keeping with such flexible approaches, in this paper, we propose a novel OpenIE-format unsupervised relation extraction method for knowledge graph generation.

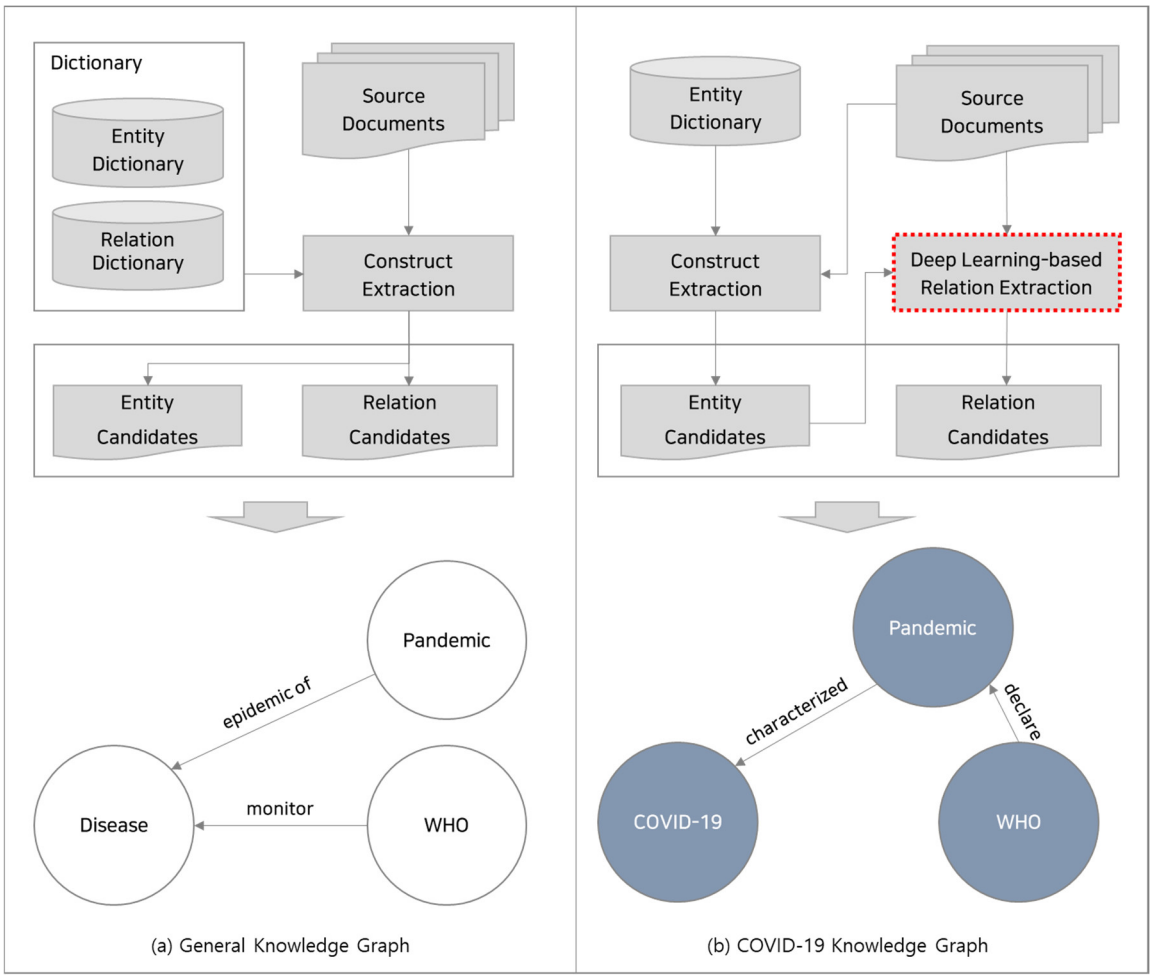

Figure 1. (**a**) Representing relationships between COVID-19-related entities using a general knowledge graph. In (**a**), the graph focuses on generally identified knowledge, such as causality or hierarchical relationships. Entities and relations are extracted based on the entity dictionary and relation dictionary, respectively. (**b**) A COVID-19-specific knowledge graph can explain COVID-19-focused relationships between the entities in the general knowledge graph. This domain-specific

knowledge graph can define novel knowledge that cannot be discovered in the general knowledge graph. The descriptions or direction of relationships that exist in the general knowledge graph can be modified.

Figure 1a shows that defining specific roles and the meaning of entities in a specific field is impossible when constructing a domain-specific knowledge graph using traditional "dictionary-based" or "supervised" knowledge graph construction methods, because entities and relations are extracted based on previously defined dictionaries. In this figure, an entity refers to the nodes of the KG, and roles refer to the relation between entities. For example, two entities, i.e., "disease" and "pandemic" are related by "epidemic of", which represents causality in Figure 1a. On the other hand, in Figure 1b the entity "pandemic" is related to new entity, i.e., "COVID-19", and these two entities are related by "characterized", which expresses an expansion of the meaning through the progress of the event. This difference occurs because, typically, the traditional "dictionary-based" or "supervised" knowledge graph defines relationships in order to represent the linguistic meaning of the entities. Therefore, when adding knowledge of a specific domain such as COVID-19 to the existing KG, applying traditional knowledge graph building methods that rely on predefined human-annotated resources is difficult because it only attempts to represent the ontological meaning of entities and will not allow rapid adaptability in describing new knowledge when domain-specific information is added or updated very frequently.

Thus, in this study, we propose a deep learning-based COVID-19 relation extraction method for knowledge graph construction that does not originate from a traditional knowledge base. Normally, a pre-built head–relation–tail (HRT)-formatted knowledge base is used to construct a knowledge graph. However, in specific fields, such as COVID-19, where relationships between various entities need to be redefined indicating relationships, using a pre-built general knowledge base is difficult. Rather, it can be more effective to extract new COVID-19-specific knowledge from COVID-19-related text and add it to a directional graph. Therefore, our model in Figure 1b attempts to dynamically extract new relations from given texts using deep learning without any dependence on pre-built relation dictionaries. Our model is not intended to replace the existing knowledge graph generation methodology, but rather offers a way to add new knowledge to the existing knowledge graph. In detail, we proposed an unsupervised learning-based OpenIE system. Our model can identify a relation between COVID-19-related entities using the attention weight of our Co-BERT model without any pre-built training dataset.

We introduce an unsupervised COVID-19-relation extraction method for updating knowledge graphs using the bidirectional encoder representations from transformer (BERT) pre-trained language model. The proposed method was developed in three steps. First, we constructed a COVID-19 entity dictionary. Then, we scraped an extremely large news and scientific paper dataset related to COVID-19. Documents were separated into sentences, and, using these sentences and the COVID-19 entity dictionary, we constructed a COVID-19-focused language model by fine-tuning the BERT. Finally, we defined a new COVID-19-specific HRT-formatted knowledge base by extracting relational words using BERT self-attention weights from COVID-19 sentences and generated this knowledge base as a directional knowledge graph. We evaluated the proposed method by measuring our fine-tuned BERT model's training accuracy and comparing the proposed method to traditional knowledge and COVID-19-focused knowledge.

The remainder of this paper is organized as follows. In Section 2, we discuss previous studies related to the proposed relation extraction method for knowledge graph generation, and, in Section 3, we explain the proposed method in detail and provide some examples. In Section 4, we present the results obtained by applying the proposed method using actual COVID-19-related data and compare the proposed method to an existing model. We summarize our contributions and provide suggestions for future work in Section 5.

2. Related Work

2.1. Pre-Trained Language Model

A language model assigns contextual probabilities to a sequence of words, and can be used to predict the probability of an entire sentence or to predict the next word given a previous sequence of words. Language model research is being actively conducted, and language models have demonstrated excellent performance in machine translation, QnA Chatbot, and knowledge graph generation, which are key areas in natural language understanding and processing. Natural language processing is divided into two methods, a statistical method and a method that uses using a neural network. Traditionally, the statistical method has been used; however, the entire statistical calculation must be repeated when a new vocabulary item is added. On the other hand, neural networks can infer new data, and a target network can be fine-tuned based on the weights of a similar, previously trained model such that the target network can be efficiently expanded. Neural networks have attracted considerable attention recently.

In particular, deep learning-based language models demonstrate state-of-the-art (SOTA) performance and have made a leap of progress owing to the concept of attention mechanism. Well-known representative deep learning-based pre-trained language models include: the embeddings from language model (ELMo), BERT, the extra long network (XLNet) [3–5], and generative pre-trained transformer 3 (GPT-3) [6]. ELMo is a bidirectional language model that uses LSTM to learn text sequences in two directions, forwards and backwards. BERT is also a bidirectional language model that has been proposed to handle the long-term dependencies problem, which is a chronic limitation of the LSTM model. It learns based on the self-attention mechanism, which recognizes the contextual relationship between a specific word and other words in the same sequence. Especially, the transformer [7] structure adopts this self-attention mechanism and accomplishes remarkable performance without using any RNN or LSTM cells. Almost all deep learning-based pre-trained language models, e.g., XLNet and GPT-3, that were introduced after BERT are variant models based on this structure. Such language models attempt to generate a model that universally understands natural language by learning the model using a huge corpus. The purpose of this type of learning is to improve the performance of natural language processing by fine-tuning downstream tasks [8–12]. This transfer learning delivers SOTA performance because it does not need to newly learn a large amount of data. Among them, BERT is widely used in various fields, and is often implemented in a specific domain such as with Bio-BERT [13] and Med-BERT [14]. Our method was also inspired by these studies. However, COVID-19 should be able to deal with COVID-19-related issues in various fields as well as in a specific field. Therefore, we extracted COVID-19-related sentences from news articles as well as scientific papers to perform fine-tuning to construct our Co-BERT model. In this paper, our Co-BERT model is implemented in two versions: Co-BERT(Vanilla), which uses general BERT as a pre-training model, and Co-BERT(Bio), which uses Bio-BERT [13] as a pre-training model.

The BERT learning method used in this study can be divided into two categories. The first is the masked language model (MLM), which is a learning method that predicts the word by putting a mask on a part of the input word. The second is the next sentence prediction (NSP) method, which predicts whether the second sentence is the sentence immediately following the first sentence when two sentences are given. Differing from the existing left–right language model learning method, which performs learning according to the sequence of sequences, MLM replaces the word in the middle of the sequence with a mask and performs learning to predict the word. In general, noise is inserted into 15% of the input words. Of that 15%, 80% are covered with a MASK token, 10% are another random word, and 10% are left as is. This increases the generalization performance of transfer learning by intentionally inserting noise in the model's learning process. The MLM is used for various purposes besides building language models [15–17], and our study aims to present a new application method based on these previous applications. On the other

hand, with NSP, 50% of the input is an actual continuous sentence, and the remaining 50% is a randomly extracted sentence. Learning is performed to predict whether the sentence is a continuous sentence or a randomly generated sentence. A language model can effectively learn general characteristics of the language and the context of sentences using both the MLM and the NSP method. In this study, using BERT's MLM model, a mask covers a pre-defined word and an entity, and the relationship between the words is determined through the self-attention weight. We define this relationship as a "triple", and use it to build a knowledge base from the perspective of COVID-19.

2.2. Knowledge Graph (KG)

A knowledge graph is a type of knowledge base that Google applied to their information retrieval system. Google used this concept to improve search results by expanding the scope of information related to input queries. For example, if a user searches a specific person, Google presents basic information about the person, such as their date of birth, job, organization, and related people. In addition, when a user enters a more detailed query, such as the place of birth or height, the search engine will return an exact answer. As a successful implementation of Google's case, the word "knowledge graph" has been considered as a representative of an existing knowledge base. Due to the increasing interest in and the improved performance of knowledge graphs, many studies related to knowledge graphs have been conducted, and the fields that use knowledge graphs are continuously increasing.

Knowledge graphs were first applied to information retrieval because they allow logical inference for retrieving implicit knowledge. However, as the artificial intelligence (AI) domain continues is to develop, and AI applications diversify, many studies have attempted to apply knowledge graphs to various AI-related domains. Knowledge graphs have been applied to chatbots and question answering systems. Chatbots are software-based online human to machine dialog systems [18]. Recent chatbot studies investigated automatic dialog generation based on human dialog input using AI technologies [19]. Representative uses of chatbot software are general messenger applications and enterprise messenger systems for general conversation. Chatbot software can also focus on domain-specific user to machine interaction systems, such as online shopping mall shopping assistant services and healthcare applications. Recent studies have used knowledge graphs to enhance chatbot performance. More specifically, there are general purpose knowledge graph-based chatbots for transferring knowledge or conducting common conversations [20,21], chatbots for item searching and questioning in online shopping services [22], chatbots that provide tourism information [23], and domain-specific chatbots based on knowledge graphs for self-diagnosis, medical system assistance, and online diagnostic assistance [24–27].

In addition, some studies have proposed new knowledge graph construction methods. These new methods are generally used to generate domain-specific knowledge graphs. These methods usually involve extracting and redefining domain knowledge from new data and representing the domain knowledge as a knowledge graph, because traditional knowledge graphs will not allow rapid adaptability in describing new relationships in terms of a specific domain. Various studies have applied these new methods to generate knowledge graphs for academic search and ranking mechanisms [28,29], user or item information-based knowledge graphs to construct recommender systems [30–32], and knowledge graphs for healthcare systems using medical information [33,34]. In addition, such new construction methods have been used to achieve specific goals, such as to construct multi-lingual knowledge bases, knowledge graphs for educational assistant systems, and knowledge graphs for claim-checking services [35–37].

Some studies have attempted to attach a knowledge graph to an AI-based training model. These studies usually involve "graph embedding", which converts entities and relationships in a graph to numeric features. Graph embedding is a type of topological graph theory whereby objects are mapped to dimensional space. Features generated with

graph embedding can reflect contextual information, such as individual nodes and link directions, and these features can be used for various calculations, e.g., measuring the distance between nodes or finding paths. Thus, for example, graph embedding has been applied to knowledge inference and used to evaluate the quality of a knowledge graph. Previous studies into knowledge graph embedding have investigated using arithmetic algorithms to generate embedded knowledge graph features [38–40], machine learning-based knowledge graph embedding [41,42], and knowledge embedding through the addition of text data [43]. Recently, some studies have explored the use of deep learning algorithms, such as convolutional neural networks and recurrent neural networks, to generate embedded knowledge graphs and apply them to the construction of question answering and recommender systems [44–46].

2.3. Pre-Trained Language Model-Based KG

Methods to construct domain-specific knowledge graphs have also been investigated. However, because constructing a knowledge graph incurs significant time and computation costs, many studies into domain-specific knowledge graphs use pre-trained language models. Pre-trained language models are trained using unsupervised learning, which does not require labeled data. Therefore, studies that propose methods to construct knowledge graphs using a pre-trained language model usually also propose adding a small amount of text data and fine-tuning the model to extract entity to entity relationships. Among various unsupervised pre-trained language models, BERT, which uses the transformer algorithm, is frequently applied to construct domain-specific knowledge graphs. When using BERT to construct a knowledge graph, it is possible to extract knowledge and relationships from new data that include unseen knowledge; therefore, this method can generate highly extensible knowledge graphs.

There are two main approaches for constructing knowledge graphs using BERT. The first approach constructs a knowledge graph by masking specific entities, predicting them correctly, and extracting them as knowledge [47]. Another approach inputs existing knowledge bases into BERT and generates a corresponding knowledge graph that can be expanded continuously [48]. In addition, there is research that emphasizes the compatibility of knowledge using HRT-formatted sentences that contain knowledge as BERT input data and then measures the score of the generated knowledge [49]. Other studies predict relational words to represent a knowledge graph or classify various pre-defined relational labels of input sentences that include knowledge to generate a knowledge graph [50,51].

Methods for knowledge graph embedding using BERT have also been proposed recently. When using BERT for knowledge graph embedding, it is possible to map contextual representations that consist of entities and relationships into fixed vector space, and then infer vectors of new knowledge that does not exist in the model training process. For example, a method that uses BERT to extract vectors of sentences that contain general knowledge has been proposed [52]. Various other studies attempt to derive special vectors that can achieve a unique goal. For example, a method to transplant a knowledge base into an intelligent chatbot using BERT-based knowledge graph embedding has been proposed [53]. Embedding an existing knowledge graph using BERT has also been applied to various text classification tasks and used to increase document classification accuracy [54,55].

2.4. Open Information Extraction (OpenIE)

Open information extraction (OpenIE) is the task of extracting information from a large-scale text corpus, usually in a form of triples of *n*-ary propositions [56]. It is useful in many NLP tasks, such as information retrieval [57], question answering [58], and relation extraction [59]. In particular, it can be useful for extracting relations without any reliance on a pre-defined schema. For example, given the sentence "pandemic is an epidemic of disease", the triple can be extracted as: pandemic; is an epidemic of; disease.

After the OpenIE system was first proposed by TextRunner [60], several systems such as ReVerb [61] and Stanford OpenIE [62] were proposed. However, most of these systems operate on a rule-based basis although they use some NLP techniques for syntactic parsing and named entity recognition. Recently, a few relation extraction methods based on machine learning have been proposed. However, most of them perform supervised learning from a pre-built training set with a triple structure [63,64]. On the contrary, in this paper, we propose an unsupervised learning-based OpenIE system. In detail, our model can identify a relation between COVID-19-related entities using BERT without any pre-built training dataset.

3. Proposed Methodology

3.1. Research Process

In this section, we explain the proposed COVID-19-specific relation extraction for the knowledge graph updating method and provide examples. The overall research process is represented in Figure 2.

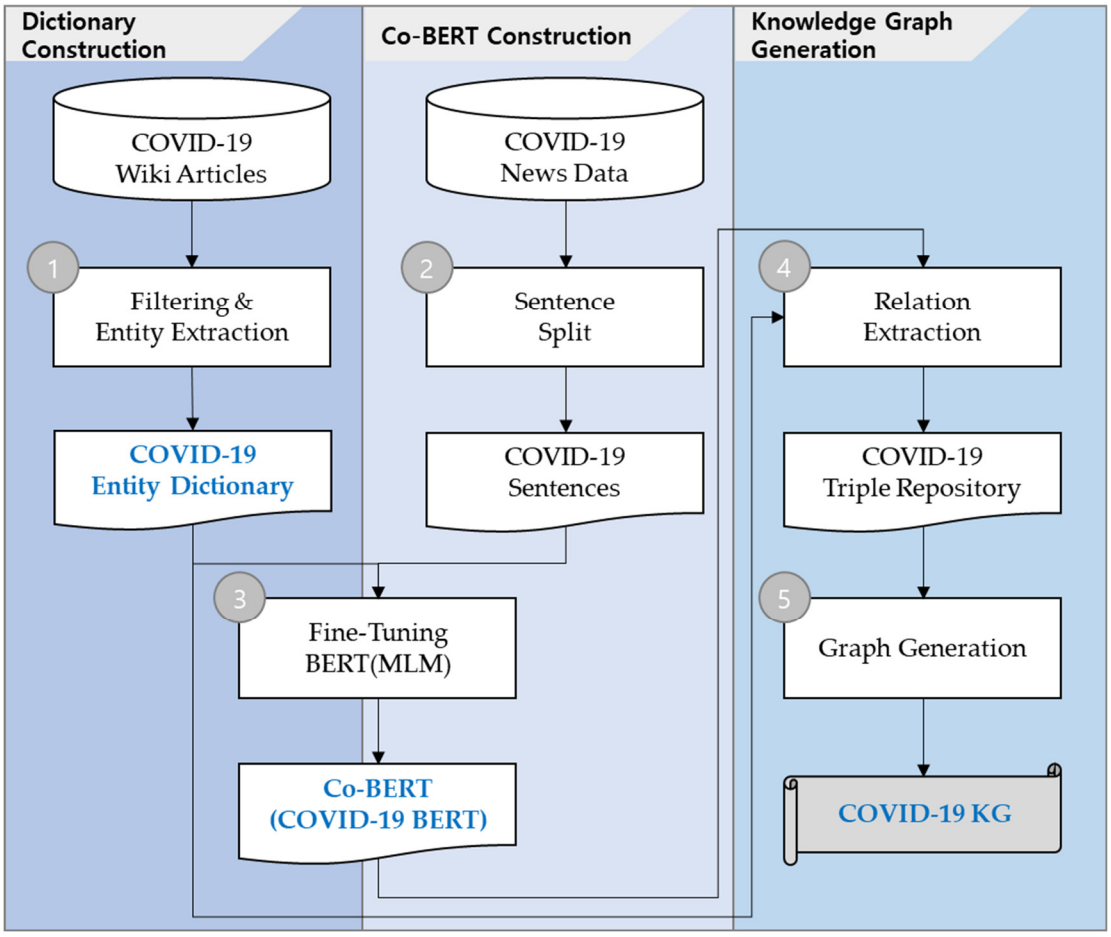

Figure 2. The overall COVID-19-specific relation extraction method for the updating of the knowledge graph. The proposed method involves three tasks: (Ⅰ) COVID-19 entity dictionary construction, (Ⅱ) Co-BERT construction using the entity dictionary and COVID-19 news or scientific papers, and (Ⅲ) HRT-formatted knowledge extraction and knowledge graph generation using Co-BERT in sequence.

As shown in Figure 2, the overall research process involves extracting relations for the COVID-19 knowledge base and generating a knowledge graph using a COVID-19-specific language model trained with COVID-19 sentences and Wikipedia articles. When constructing the Co-BERT model, we used news articles as a new option to quickly and

automatically extract general knowledge about rapidly changing issues in various domains, and scientific papers as another option to extract knowledge from the scientific perspective. However, since there is no significant difference in the detailed methodology regarding the data, further description will be made mainly based on news data for convenience of explanation. The COVID-19 entity dictionary construction task involves extracting useful entities from COVID-19 Wikipedia articles. The COVID-19-specific language model (Co-BERT) is generated from COVID-19 news, and the entity dictionary and knowledge graph are generated using COVID-19 knowledge extraction using the proposed Co-BERT. Specifically, in the dictionary construction task, we extract all candidate items that can be entities from English Wikipedia COVID-19-related articles and construct the COVID-19 entity dictionary by applying original filters that we defined to all candidate items. In the Co-BERT construction task, we split the COVID-19 news into sentences and fine-tuned them using the COVID-19 entity dictionary using a BERT MLM to construct the COVID-19-specific language model. Finally, to generate the knowledge graph, we constructed a COVID-19 triple formatted knowledge repository by extracting relational words that can explain relationships between all pairs of entities using Co-BERT and represent the entities and their relationships as a directional knowledge graph.

In the following sections, the detailed process of each task will be explained, and hypothetical examples will be provided. In Section 4, we will present the results obtained by applying the proposed method applied to COVID-19 news and Wikipedia articles, and evaluate the results.

3.2. COVID-19 Dictionary Construction

In this section, we introduce the method used to construct an entity dictionary that stores various entities that can be nodes of the COVID-19-specific knowledge graph. Various methods can be used to construct dictionaries, such as using an existing corpus, e.g., WordNet, or part of speech-tagging or syntax analysis to extract terms that have unique meanings. For a distinct platform, like Wikipedia, information, such as categories, template types, number of searches, or number of views, are used when extracting available terms. In this study, we constructed the COVID-19 entity dictionary by defining various filters, such as filters based on categories or information content, and applying the filters to all candidate item groups. To construct a COVID-19 entity dictionary, first we extract all candidate items from coronavirus 2019 portal articles in the English Wikipedia [65]. Specifically, we extract available items from all articles that have a hyperlink to other articles. After collecting all candidate items, we define filtering rules to remove general terms or items not related to COVID-19, and apply these rules to all candidate entities to construct the COVID-19 entity dictionary. Table 1 shows some examples of the filtering rules used in this study.

Table 1. We defined some filters to elaborate the COVID-19 entity dictionary by excluding some noisy items. Entity terms must deliver one of the smallest units of meaning, so terms that combine multiple meanings should be removed. For example, "COVID-19 pandemic in the United States" should be removed from the dictionary because it cannot be regarded as one of the smallest units of meaning.

Filter ID	Filters	Defined by	Filter ID	Filters	Defined by
Filter 1	COVID-19 pandemic ~	Title	Filter 6	Company	Category
Filter 2	by ~	Title	Filter 7	Encoding Error	Contents
Filter 3	Impact of ~	Title	Filter 8	President of ~	Category
Filter 4	list of ~	Title	Filter 9	lockdown in (place)	Title
Filter 5	pandemic in (place)	Title	Filter 10	External Link	Content

We use item content and category information to define filters and apply them to all candidate items. For example, in Table 1, Filter 1 and Filter 2 are defined from item titles,

and Filter 6 and Filter 8 are defined from an item's category information described in the bottom of Wikipedia articles. Only items not removed by the filtering process are stored in the COVID-19 entity dictionary, and this dictionary will be used in the Co-BERT construction and knowledge graph generation.

3.3. Co-BERT Construction

In this step, a new language model from the perspective of COVID-19 is constructed by fine-tuning BERT's MLM using the constructed COVID-19 entity dictionary. When learning MLM, the internal BERT model, a new method that involves masking only the entities included in the COVID-19 dictionary is used rather than the traditional method of random masking. The MLM predicts words that will enter MASK by grasping the relationship and context between surrounding words in a sentence. Through the attention mechanism, the masked entity and words that are closely related in terms of context are learned to receive high attention weight, and the relationship between entities can be inferred. In other words, constructing Co-BERT with fine-tuning that predicts entities using MLM is an essential process to extract specialized knowledge. In addition, as the performance of Co-BERT increases, the knowledge of the HRT structure can be clearly expressed from the perspective of Covid-19. Therefore, this study aims to guarantee the excellence of the generated knowledge base by enhancing the performance of Co-BERT.

Figure 3 shows the learning process of our methodology. Co-BERT construction is divided into three levels: tokenizing, masking, and fine-tuning. First, the input sentence "COVID-19 originated in Wuhan, Hubei" is divided into tokens using the Co-BERT tokenizer. Next, a Co-BERT model is constructed by putting MASK on "Covid-19" and "Wuhan", corresponding to entities among the tokens of the divided sentence and performing learning to predict them.

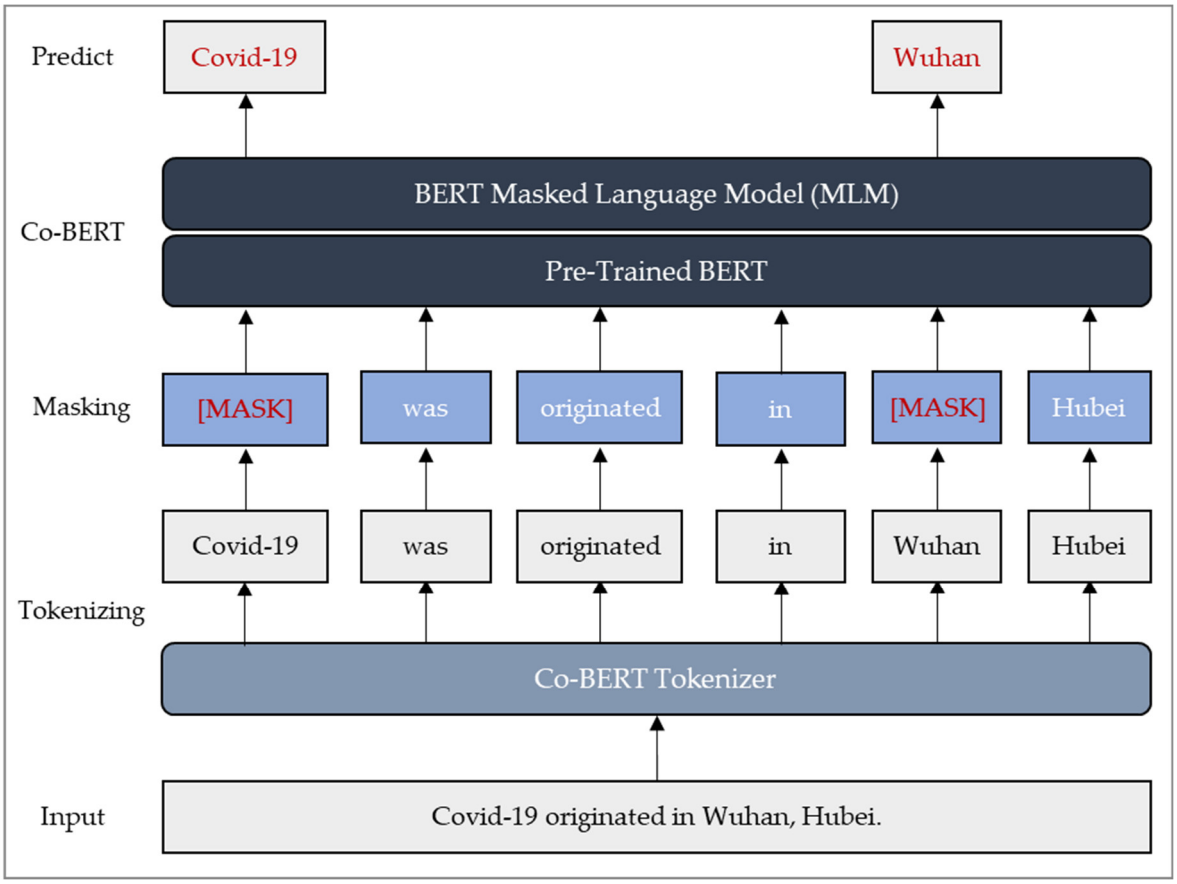

Figure 3. Diagrammatic representation of the Co-BERT model process from the input to output. After the original sentence is inputted, the process masks a specific entity and predicts the masked tokens.

In Figure 3, the masked sentence is converted to a number that can be used to learn text through the BERT tokenizer. The existing BERT tokenizer consists of 30,522 tokens that are often not properly identified as they do not contain new words related to COVID-19. For example, the entity "COVID-19" used in the input sentence is not properly identified by the general BERT tokenizer, and is divided into "Co", "VID", "-", and "19." Given that many of the terms related to COVID-19 are new words, these issues must be addressed to generate a knowledge graph related to COVID-19. Therefore, in this study, the Co-BERT tokenizer was built by adding words in the COVID-19 dictionary to the BERT tokenizer in order to perform tokenization that reflects COVID-19-related entities.

First, the COVID-19 entity dictionary is used in the process of masking the input sentence. When a specific token of the input sentence is included in the COVID-19 entity dictionary, masking is performed, and the number of entities subject to masking is limited to two per sentence. The entity that appears first in the sentence is the head entity, and the entity that appears later is the tail entity, and the relationship between the two entities is learned. In the input sentence in Figure 3, it can be seen that the words "COVID-19" and "Wuhan" are included in the COVID-19 dictionary, and each has a mask.

The masking process can be classified into four types according to the number of Covid-19-related entities in the input sentence. The four types are summarized in Table 2. If there are only two entities in the sentence, masking can be performed on the two entities directly; however, in other cases, there are several methods. If the number of entities in a certain sentence is 0, we can exclude the sentence from the training data or we can randomly select and mask two tokens. When a sentence contains one entity, we can also exclude the sentence from the training data or we can randomly select and mask one entity additionally. Even when the number of entities is three or more, there are two cases where only two of the entities are randomly selected and excluded from the training data.

As described above, the masking method can be applied differently depending on the number of entities included in the sentence. In Table 2, Version 1 is a method used as training data only when the number of entities included in the sentence is two, which ensures that only sentences that can represent the specific perspective of COVID-19 in learning are used. This method can reflect the perspective of COVID-19 in a direct way. In addition, since there are only two entities from the perspective of COVID-19 in the sentence, it is most suitable for HRT triple knowledge extraction. Therefore, in this paper, we primarily introduce experiments on Version 1 with these characteristics.

Table 2. The method of masking input sentences is divided into four cases (where a COVID-19 entity is not included, 1 COVID-19 entity is included, 2 COVID-19 entities are included, or 3 or more COVID-19 entities are included), and four versions of the experiment in which masking is performed in different ways for each case are presented.

Version	Entity = 0	Entity = 1	Entity = 2	Entity = 3
Version1	Excluded	Excluded	2 Entities	Excluding
Version2	Excluded	Excluded	2 Entities	2 Random Entities
Version3	Excluded	1 Random Token & 1 Entity	2 Entities	2 Random Entities
Version4	2 Random Tokens	1 Random Token & 1 Entity	2 Entities	2 Random Entities

Sentences that are masked and tokenized pass through the BERT layer and the MLM layer (Figure 3) in turn and learn to predict the words that will enter MASK. The BERT layer uses pre-trained weights, and the weights are updated in the MLM layer through fine-tuning.

3.4. COVID-19-Based Knowledge Generation

COVID-19-based knowledge generation is a process of constructing a knowledge system from the perspective of COVID-19 using the Co-BERT model described in Section 3.3. Specifically, the relation between the head entity and the tail entity is extracted and a

knowledge graph of the HRT structure is generated. Twelve transformer layers are used for BERT, and twelve heads exist in each layer. The BERT attention weight, which represents the relationship between words in a sentence, is learned in each of the 12 layers and 12 attention heads. Since each learns a different context, the weights appear different depending on the viewpoint [66–70]. This is a major cause of difficulties in interpreting attention weight in many studies using BERT. To solve this problem, several studies have proposed a new application of the weight of BERT [71–73]. In this study, we present a new method of utilizing the BERT attention weight and propose a method of generating a knowledge graph.

The higher the attention weight learned in each layer and head, the stronger the compatibility between words in the sentence. Figure 4 is an intermediate product of Co-BERT learning and is an example of an attention map visualizing the attention weights. In the attention map, a row represents a criterion word, and a column represents a word related to the criterion word. Note that 144 attention maps can be obtained for each layer and head.

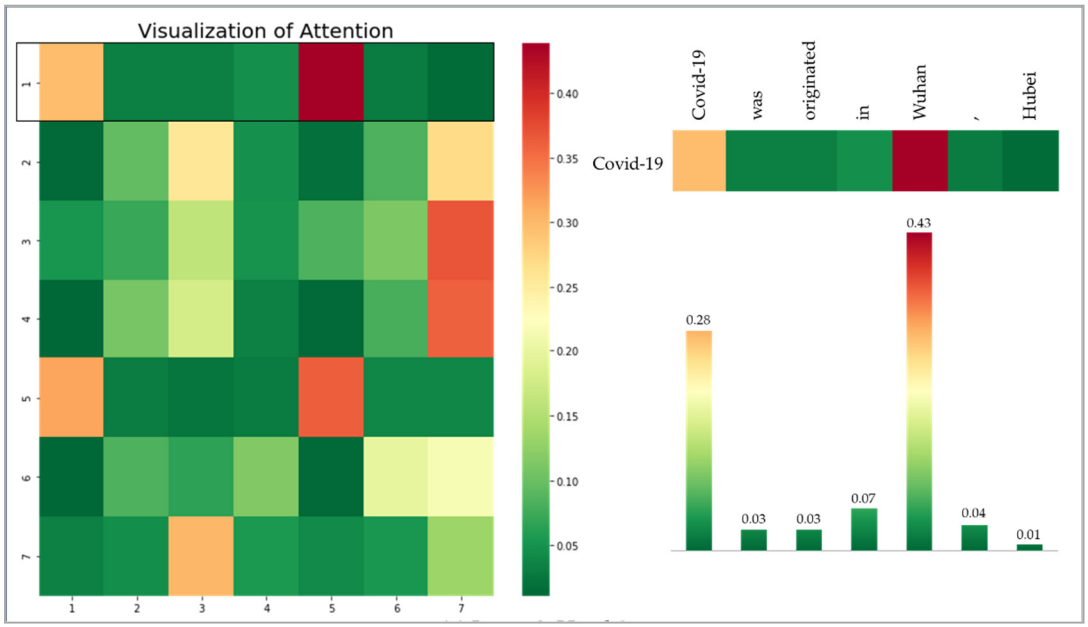

Figure 4. Visualization of the attention weights between words and words included in the sentence, "Covid-19 originated in Wuhan, Hubei". The closer the color to red, the higher the weight. For example, the red cell in the left matrix (1, 5) is indicating that the word most related to "Covid-19" is "Wuhan".

We present a method to utilize the association between words that can be obtained through the attention map. In other words, a method to extract the relation between the head entity and tail entity using the attention weight is proposed. To this end, the word most related to the two entities is searched for in all attention maps, and the number of occurrences is counted for each word. For example, in Figure 4b, the head entity is "Covid-19", which is the most correlated with "originated", and the tail entity "Wuhan" is also the most correlated with "originated." Here, if the number of words with the highest correlation with entities for 144 attention maps is counted, an aggregate value (Table 3) can be calculated for each of the two entities. For example, as shown in Table 3, the value of the cell corresponding to "Covid-19" and "originated" is 12, which means that, among the 144 attention maps for predictions masking "Covid-19" and "Wuhan", there are a 12 maps with the highest attention weight of "originated" from the perspective of "Covid-19".

Table 3. The relationship between each word in the target sentence is shown.

	Covid-19 (Head Entity)	Wuhan (Tail Entity)	Covid-19 * Wuhan
Covid-19	11	4	44
was	7	0	0
originated	12	9	108
in	1	6	6
Wuhan	4	9	36
Hubei	2	15	30

The operator(*) in the 4th column means multiplication of head entity and tail entity.

Words that can define the relationship between two entities were selected by calculating element-wise production for the aggregated values of each of the two entities. The fourth column in Table 3 shows the element-wise production value, and the word with the highest value is defined as the relation of two entity combinations. In other words, the relation between "Covid-19" and "Wuhan" is defined as "originated" (with the highest value of 108). However, in this process, the two words that were the target of masking were excluded from the relation candidates. In addition, special tokens attached to the front and back of the sequence were excluded from relation candidates. The proposed method can construct knowledge from the perspective of COVID-19 with a head–relation–tail (HRT) structure using the attention weight between words through the above process.

4. Experiment

4.1. Overall

In this section, we evaluate the experimental results obtained by the proposed method using real data. To implement this experiment, we scraped COVID-19 news data from 01 January 2020 to 30 June 2020 and candidate items from the coronavirus disease 2019 portal in English Wikipedia. The search queries were "COVID-19", "coronavirus", "pandemic", etc. Finally, we used approximately 100,000 sentences from news data and a filtered entity dictionary. The experimental environment was constructed using the PyTorch deep learning framework in Python 3.7.

We also conducted experiments on scientific papers in consideration of the medical and scientific nature of COVID-19. We used data from the "COVID-19 Open Research Dataset Challenge (CORD-19)", which is a scientific paper dataset on COVID-19 publicly available in kaggle. From 400,000 abstracts of the papers, approximately 100,000 COVID-19-related sentences were randomly selected and used for training. Unlike news from a general and popular viewpoint, scientific papers were written from the viewpoint of a specialized domain, so experiments were conducted using Bio-BERT [13] as well as Vanilla BERT for fine-tuning. The performance comparison of Co-BERT(Vanilla), a version using Vanilla BERT for pre-training, and Co-BERT(Bio), a version using Bio-BERT for pre-training, is made in the evaluation section.

4.2. Constructing the COVID-19 Dictionary

Here, we present the results of extracting candidate items from Wikipedia's coronavirus disease 2019 portal and constructing the entity dictionary with COVID-19-related items extracted from all candidate items using filters. We extracted 24,845 candidate items from Wikipedia articles in the coronavirus portal with a hyperlink to other Wikipedia pages. Then, we defined several filters to extract the COVID-19-related entities used in a knowledge graph and selected valid entities by applying the defined filters to all items. Here, 12,196 entities were extracted from the 24,845 items, and these entities were used to construct the COVID-19 entity dictionary. Table 4 shows examples of the filtered entities.

Table 4. Samples of the entity dictionary filtered from the Wikipedia coronavirus disease 2019 portal items.

No.	Entities	No.	Entities	No.	Entities	No.	Entities
1	Pandemic	11	Hand washing	21	Zoonosis	31	Herd immunity
2	Wuhan	12	Social distancing	22	Pangolin	32	Epidemic curve
3	Fatigue	13	Self-isolation	23	Guano	33	Semi log plot
4	Anosmia	14	COVID-19 testing	24	Wet market	34	Logarithmic scale
5	Pneumonia	15	Contact tracing	25	COVID-19	35	Cardiovascular disease
6	Incubation period	16	Coronavirus recession	26	Septic Shock	36	Diabetes
7	COVID-19 vaccine	17	Great Depression	27	Gangelt	37	Nursing home
8	Symptomatic treatment	18	Panic buying	28	Blood donor	38	Elective surgery
9	Supportive therapy	19	Chinese people	29	Cohort study	39	Imperial College
10	Preventive healthcare	20	Disease cluster	30	Confidence interval	40	Neil Ferguson

4.3. Constructing Co-BERT

Here, we present the results of constructing the Co-BERT pre-learning language model from the perspective of COVID-19 using the collected news and COVID-19 dictionary. The collected 100,000 COVID-19-related news articles which were divided into sentences to form a data pool of approximately 200,000. By performing the Version 1 masking shown in Table 2, 15,000 training datasets were constructed. The COVID-19-specific training dataset was used to fine-tune the construction of the final Co-BERT model.

Co-BERT learned a perspective specific to the COVID-19 situation. This is confirmed from the mask prediction result obtained using Co-BERT. Figure 5 shows the results of predicting masks in each sentence by the general BERT and Co-BERT models. As shown, the general BERT completed the sentence from a general perspective; however, Co-BERT clearly predicted the sentence from a COVID-19 perspective.

No.	Model	Input & Predicted Output
1	Input	[MASK] was originated in [MASK] , hubei
	General Model	**It** was originated in **Germany** , hubei
	Co-Bert Model	**Covid-19** was originated in **wuhan** , hubei
2	Input	The [MASK] cause severe illness, such as [MASK]
	General Model	The **fungus** cause severe illness, such as **cancer**
	Co-Bert Model	The **coronavirus** cause severe illness, such as **pneumonia**
3	Input	[MASK] was first identified in [MASK] in December 2019
	General Model	**It** was first identified in **science** in December 2019
	Co-Bert Model	**Covid-19** was first identified in **china** in December 2019
4	Input	[MASK] declares [MASK] as public enemy number one
	General Model	**He** declares **himself** as public enemy number one
	Co-Bert Model	**China** declares **coronavirus** as public enemy number one
5	Input	[MASK] at home has been recommended for those diagnosed with [MASK]
	General Model	**Treatment** at home has been recommended for those diagnosed with **cancer**
	Co-Bert Model	**Quarantine** at home has been recommended for those diagnosed with **covid-19**

Figure 5. Mask prediction results for five input sentences. These sentences were not included in the training set and were only used for the test. The general model is the inference result of the BERT model reflecting the general perspective. The Co-BERT model is the inference result of the model reflecting a COVID-19 perspective by applying the proposed method. The highlighted part is the result of predicting the word.

4.4. Generating COVID-19-Based Knowledge

This section presents the results of extracting the word representing the relation between the head entity and tail entity using the Co-BERT model, and a part of the result of generating the extracted knowledge base specialized for COVID-19 as a knowledge graph. The extracted knowledge base and visualized knowledge graph are shown in Table 5 and Figure 6.

Table 5. Some knowledge triples specific to COVID-19 derived through the Co-BERT model are shown.

No	Head Entity	Relation	Tail Entity
1	People	died	Disease
2	Disease	called	COVID-19
3	COVID-19	confirmed	Nigeria
4	China	lockdown	Wuhan
5	Wuhan	suspended	Public Transportation
6	People	leaving	Wuhan
7	Pneumonia	caused by	COVID-19
8	WHO	declare	Pandemic
9	COVID-19	characterized	Pandemic
10	Pneumonia	is	Inflammation
11	People	sharing	False Information
12	Disease	originated	Wuhan
13	COVID-19	reported	China
14	Nigeria	control	Disease
15	WHO	recognized	Pathogen

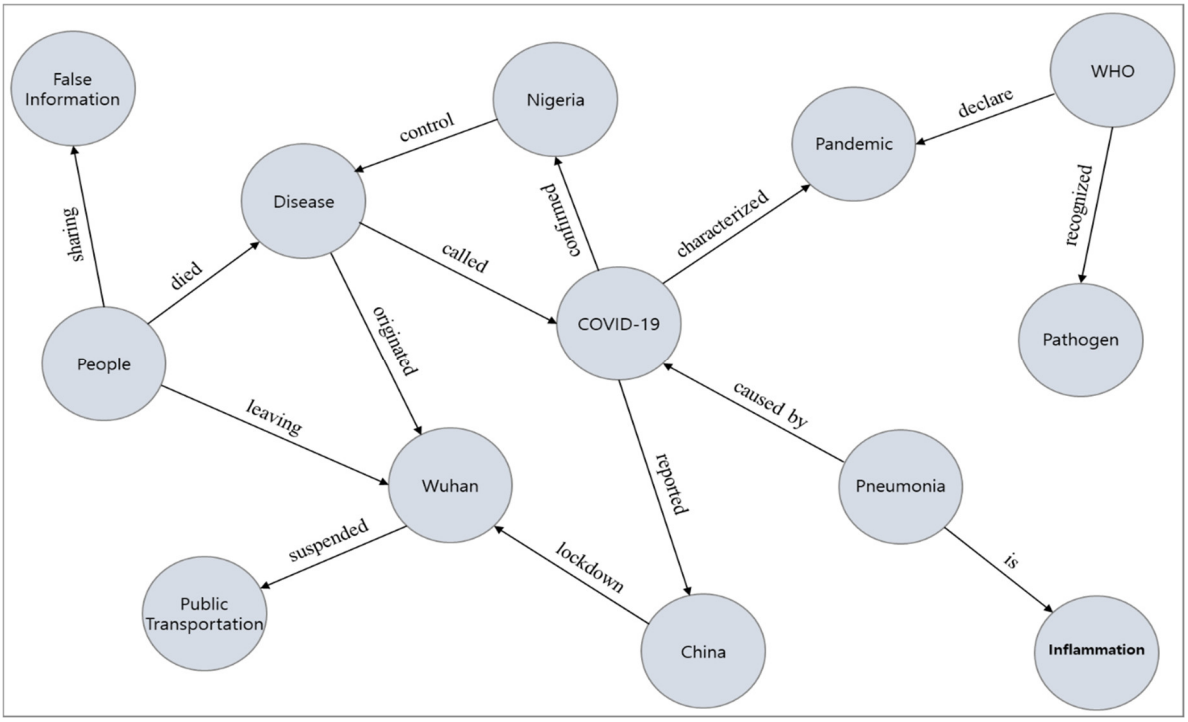

Figure 6. Part of the COVID-19 knowledge system (Table 5) as a directed graph.

4.5. Evaluation

Co-BERT

In this section, we evaluate the Co-BERT model's performance evaluation result, which was used to construct the knowledge graph. It is hard to conduct quantitative experiments on whether knowledge was well extracted. Instead, we attempt to indirectly show that the model works well by analyzing how accurately the masked entities were created from a COVID-19 perspective rather than the general perspective.

Here, we measured first the average mask prediction accuracy using pure BERT on general text. In addition, we performed the same task with COVID-19-related text and

identified reduced prediction accuracy. Finally, we measured the Co-BERT model's average prediction accuracy on COVID-19-related text and observed increased model performance compared to the pure BERT model. In each task, we observed each model's prediction accuracy after masking two words from sentences for various perspective evaluations, to establish three measuring standards. We have three evaluation criteria. The first criterion scores 1 point when the model predicts all two words accurately (tight criterion). The second scores 0.5 points when the model predicts an only a single word accurately (in between). Finally, the last evaluation criterion scores 1 point when the model predicts at least one word accurately (loose criterion).

The metric for evaluation of translation with explicit ordering (METEOR) score is a metric used in machine translation or text generation applications. The METEOR score evaluates model performance by calculating the weighted harmonic mean with precision and recall from unigram matching rates between the original text and the generated text. In addition, the METEOR score considers linguistic characteristics like contextual similarities or word-to-word relationships simultaneously. Here, we used the METEOR score to evaluate semantic similarity between the original text and the generated text from mask prediction. We used about 2000 COVID-19-related texts and 10,000 general texts for this evaluation. The results are summarized in Figure 7; Figure 8 with prediction accuracy.

Figure 7a compares the accuracy of predicting general sentences and COVID-19-related sentences with a pure BERT model. In this experiment, the pure BERT model shows significantly lower accuracy in predicting COVID-19-related sentences than in predicting general sentences, implying the need for a domain-specific model to understand COVID-19-related sentences more properly. Additionally, Figure 7b shows the experimental results predicting COVID-19-related sentences with various models. In this experiment, we compare the accuracy of three models such as a pure BERT model without fine-tuning, a Co-BERT(Vanilla) model that was fine-tuned with the COVID-19-related sentences from news articles after pre-training with Vanilla BERT, and a Co-BERT(Bio) model that was fine-tuned with the COVID-19-related sentences from news articles after pre-training with Bio-BERT. As a result of the experiment, the accuracy of the pure BERT model was lower than those of the Co-BERT models, and the accuracy of Co-BERT(Vanilla) and Co-BERT(Bio) do not show a significant difference. Similarly, Figure 8 compares the prediction accuracy when two models of Co-BERT(Vanilla) and Co-BERT(Bio) are applied to scientific papers. In this experiment, unlike the experiment using news data (Figure 7b), the Co-BERT(Bio) shows a noticeably higher accuracy than that of Co-BERT(Vanilla).

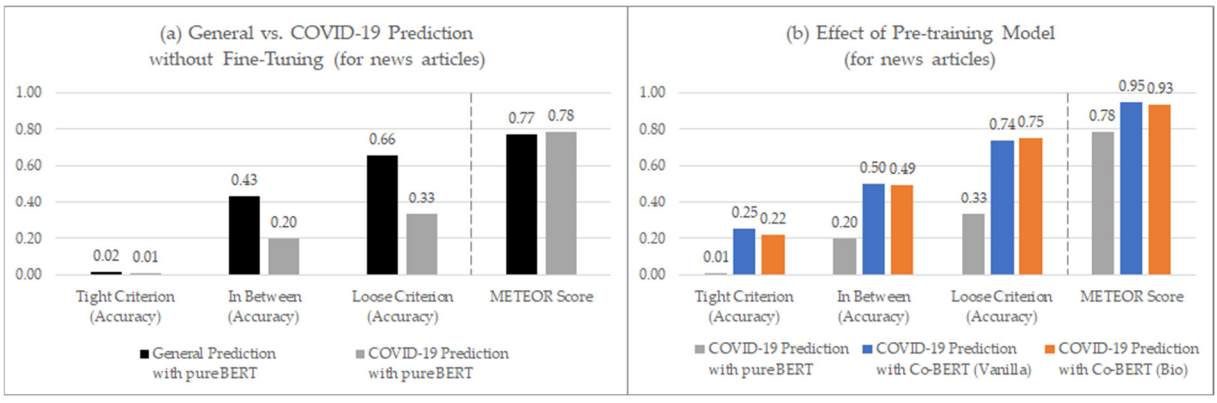

Figure 7. Visualized results of mask prediction accuracy and METEOR scores. Graph (**a**) shows that pure BERT appears to be suitable for general news, but not for COVID-19-related news. Graph (**b**) reveals that the two Co-BERT models (Vanilla and Bio) show higher accuracy than the pure BERT.

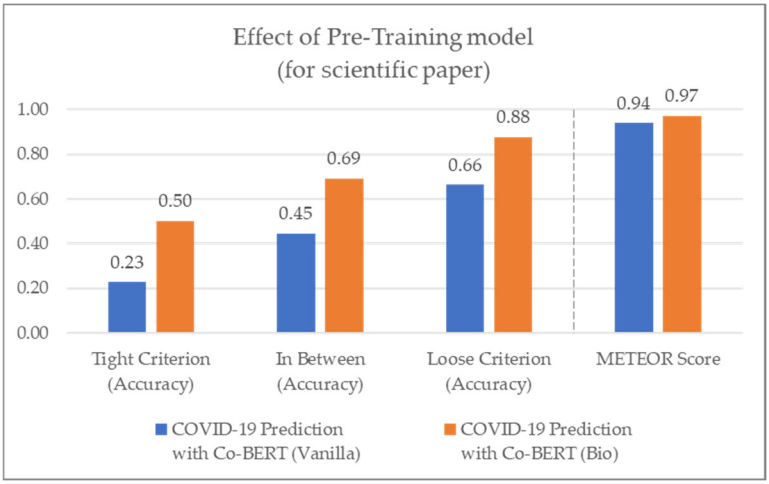

Figure 8. Visualized results of mask prediction accuracy and METEOR scores using COVID-19 scientific papers. This is the result of a comparative experiment between the Co-BERT(Vanilla) and Co-BERT(Bio) models for scientific papers. This reveals that Co-BERT(Bio) shows a higher accuracy than Co-BERT(Vanilla).

5. Conclusions

In this paper, we proposed a deep learning-based relation extraction method for COVID-19 knowledge graph generation. Specifically, we proposed an unsupervised learning-based OpenIE system which can identify a relation between COVID-19-related entities using BERT without any pre-built training dataset. In situations where information is added or updated rapidly, e.g., with COVID-19, a new knowledge system that flexibly reflects the rapidly changing situation is required. We constructed an entity dictionary from a COVID-19 perspective, and then we extracted relations between COVID-19 entities using the attention weight of our Co-BERT model.

The target topic that our Co-BERT model attempts to describe is COVID-19. However, COVID-19 is not a topic that belongs to one specific domain, but rather is a topic that is entangled in various domains. Therefore, we need to extract the relations related to the COVID-19 entities from general documents such as news articles that contain knowledge of COVID-19 from various perspectives, not from technical or scientific documents containing specialized knowledge of any specific domain [74]. News articles are also far more advantageous than scientific articles in terms of timeliness. Additionally, we utilized Wikipedia when constructing the entity dictionary because it requires relatively new words due to the character of the COVID-19 issue. However, we have not been able to conduct quantitative experiments on whether the knowledge was well extracted. Instead, we indirectly showed that the model works well by experimenting with how accurately the masked entities were created from a COVID-19 perspective rather than a general perspective by analyzing METEOR scores.

Co-BERT is expected to contribute to the practical aspect of maintaining knowledge in rapidly changing circumstance. In future, we expect that the proposed method will be applied to constructing specialized knowledge bases in various fields from different perspectives. In the current study, an experiment was conducted using only one of several methods to construct an entity dictionary and extract the relations. Thus, in future, it will be necessary to conduct experiments on various entity dictionary construction methods. In addition, it is necessary to elaborate on the relation extraction method using attention.

Author Contributions: Conceptualization, T.K.; methodology, T.K.; software, T.K., Y.Y.; validation, Y.Y.; formal analysis, T.K.; visualization, Y.Y.; investigation, T.K.; data collecting, Y.Y.; supervision, N.K.; writing-original draft preparation, T.K., Y.Y., N.K.; writing-review and editing, T.K., N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Edward, W.S. Course Modularization Applied: The Interface System and Its Implications for Sequence Control and Data Analysis; Association for the Development of Instructional Systems (ADIS): Chicago, IL, USA, 1972.
2. Google Official Blog. Introducing the Knowledge Graph: Things, Not Strings. Available online: https://blog.google/products/search/introducing-knowledge-graph-things-not (accessed on 15 December 2020).
3. Matthew, E.P.; Mark, N.; Mohit, L.; Matt, G.; Christopher, C.; Kenton, Lee.; Luke, Z. Deep Contextualized Word Representations, arXiv **2018**, arXiv:1802.05365.
4. Jacob, D.; Ming-Wei, C.; Kenton, L..; Kristina, T. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, arXiv **2018**, arXiv:1810.04805.
5. Zhilin, Y.; Zihang, D.; Yiming, Y.; Jaime, C.; Russ, R.S.; Quoc, V.L. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 8–14 December 2019; pp. 5753–5763.
6. Tom, B.B.; Benjamin, M.; Nick, R.; Melanie, S.; Jared, K.; Prafulla, D.; Arvind, N.; Pranav, S.; Girish, S.; Amanda, A.; et al. Language Models are Few-Shot Learners, arXiv **2020**, arXiv:2005.14165.
7. Ashish, V.; Noam, S.; Niki, P.; Jakob, U.; Llion, J.; Aidan, N.G.; Łukasz, K.; Illia, P. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
8. Haoyu, Z.; Jianjun, X.; Ji, W. Pretraining-Based Natural Language Generation for Text Summarization, arXiv **2019**, arXiv:1902.09243.
9. Wei, Y.; Yuqing, X.; Luchen, T.; Kun, X.; Ming, L.; Jimmy, L. Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering, arXiv **2019**, arXiv:1904.06652.
10. Ashutosh, A.; Achyudh, R.; Raphael, T.; Jimmy, L. DocBERT: BERT for Document Classification, arXiv **2019**, arXiv:1904.08398.
11. Yen-Chun, C.; Zhe, G.; Yu, C.; Jingzhou, L.; Jingjing, L. Distilling Knowledge Learned in BERT for Text Generation, arXiv **2019**, arXiv:1911.03829.
12. Jinhua, Z.; Yingce, X.; Lijun, W.; Di, H.; Tao, Q.; Wengang, Z.; Houqiang, L.; Tie0Yan, L. Incorporating BERT into Neural Machine Translation, arXiv **2020**, arXiv:2002.06823.
13. Jinhyuk, L.; Wonjin, Y.; Sungdong, K.; Donghyeon, K.; Sunkyu, K.; Chanho, S.; Jaewoo, K. BioBERT: A pre-trained biomedical language representation model for biomedical text mining, Bioinformatics **2020**, 36, 1234–1240.
14. Laila, R.; Yang, X.; Ziqian, X.; Cui, Tao.; Degui,Z. Med-BERT: Pre-Trained Contextualized Embeddings on Large-Scale Structured Electronic Health Records for Disease Prediction, arXiv **2020**, arXiv:2005.12833.
15. Marjan, G.; Omer, L.; Yinhan, L.; Luke, Z. Mask-Predict: Parallel Decoding of Conditional Masked Language Models, arXiv **2019**, arXiv:1904.09324.
16. Kaitao, S.; Xu, T.; Tao, Q.; Jianfeng, L.; Tie-Yan, L. MASS: Masked Sequence to Sequence Pre-training for Language Generation, arXiv **2019**, arXiv:1905.02450.
17. Xing, W.; Tao, Z.; Liangjun, Z.; Jizhong, H.; Songlin, H. "Mask and Infill": Applying Masked Language Model to Sentiment Transfer, arXiv **2019**, arXiv:1908.08039.
18. Weizenbaum, J. ELIAZ-A Computer Program for the Study of Natural Language Communication between Man and Machine, Computational. Linguistics **1966**, 9, 36–45.
19. Richard, C. Deep Learning Based Chatbot Models, arXiv **2019**, arXiv:1908.0835.
20. Ait-Mlouk, A.; Jiang, L. Kbot: A Knowledge Graph Based ChatBot for Natural Language Understanding over Linked Data. IEEE Access **2020**, 8, 149220–149230.
21. Kondylakis, H.; Tsirigotakis, D.; Fragkiadakis, G.; Panteri, E.; Papadakis, A.; Fragkakis, A.; Tzagkarakis, E.; Rallis, I.; Saridakis, Z.; Trampas, A.; et al. R2D2: A Dbpedia Chatbot Using Triple-Pattern Like Queries. Algorithms **2020**, 13, 217–229.
22. Shuangyong, S.; Chao, W.; Haiqing, C. Knowledge Based High-Frequency Quetstion Answering in Alime Chat. In Proceedings of the 18th International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019.
23. Sano, A.V.D.; Imanuel, T.D.; Calista, M.I.; Nindito, H.; Condrobimo, A.R. The Application of AGNES Algorithm to Optimize Knowledge Base for Tourism Chatbot. In Proceedings of the 2018 International Conference on Information Management and Technology, Jakarta, Indonesia, 3–5 September 2018; pp. 65–68.
24. Belfin, R.V.; Shobana, A.J.; Megha, M.; Ashly, A.M.; Blessy, B. A Graph Based Chatbot for Cancer Patients. In Proceedings of the 2019 5th Conference on Advanced Computing & Communication Systems, Tamil Nadu, India, 15–16 March 2019; pp. 717–721.
25. Bo, L.; Luo, W.; Li, Z.; Yang, X.; Zhang, H.; Zheng, D. A Knowledge Graph Based Health Assistant. In Proceedings of the AI for Social Good Workshop at Neural IPS, Vancouver, Canada, 14 December 2019.

26. Divya, S.; Indumathi, V.; Ishwarya, S.; Priyasankari, M.; Kalpana, D.S. A Self-Diagnosis Medical Chatbot Using Artificial Intelligence. J. Web Develop. Web Design. **2018**, 3, 1–7.
27. Bao, Q.; Ni, L.; Liu, J.; HHH: An Online Medical Chatbot System based on Knowledge Graph and Hierarchical Bi-Directional Attention. In Proceedings of the Australasian Computer Science Week 2020, Melbourne, Australia, 4–6 February 2020.
28. Xiong, C.; Power, R.; Callan, J.; Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding. In Proceedings of the 2017 International World Wide Web Conference, Perth, Australia, 3–7 April 2017; pp. 1271–1279.
29. Ruijie, W.; Yuchen, Y.; Jialu, W.; Yuting, J.; Ye, Z.; Weinan, Z.; Xinbing, W. AceKG: A Lagre-scale Knowledge Graph for Academic Data Mining. In Proceedings of the Conference on Information and Knowledge Manangement 2018, Torino, Italy, 22–26 October 2018; pp. 1487–1490.
30. Xiang, W.; Xiangnan, H.; Yixin, C.; Meng, L.; Tat-Seng, C. KGAT: Knowledge Graph Attention Network for Recommendation. In Proceedings of the Knowledge Discovery and Data Mining 2019, Anchorage, USA, 4–8 August 2019; pp. 950–958.
31. Hongwei, W.; Fuzheng, Z.; Miao, Z.; Wenjie, L.; Xing, X.; Minyi, G. Multi-Task Feature Learning for Knowledge Graph Enhanced Recommendation. In Proceedings of the 2019 World Wide Web Conference, San Francisco, USA, 13–17 May 2019; pp. 2000–2010.
32. Hongwei, W.; Fuzheng, Z.; Jialin, W.; Miao, Z.; Wenjie, L.; Xing, X.; Minyi, G. RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. In Proceedings of the Conference on Information and Knowledge Manangement 2018, Torino, Italy, 22–26 October 2018; pp. 417–426.
33. Khalid, M.M.; Madan, K.; Mazen, A.; Maqbool, H.; Fakhare, A.; Ghaus, M. Automated Domain-Specific Healthcare Knowledge Graph Curation Framework: Subarachnoid Hemorrhage as Phenotype. Expert Syst. Appl. **2020**, 145, 1–15.
34. Rotmensch, M.; Halpern, Y.; Tlimat, A.; Horng, S.; Sontag, D. Learning a Health Knowledge Graph from Electronic Medical Records. Nature Sci. Rep. **2017**, 7, 1–11.
35. Penghe, C.; Yu, L.; Vincent, W.Z.; Xiyang, C.; Boda, Y. KnowEdu: A System to Construct Knowledge Graph for Education. IEEE Access. **2018**, 6, 31553–31563.
36. Zhichun, W.; Qingsong, L.; Xiaohan, L.; Yu, Z. Cross-Lingual Knowledge Graph Alignment via Graph Convolutional Networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 349–357.
37. Andon, T.; Fafalios, P.; Boland, K.; Gasquet, M.; Zloch, M.; Zapilko, B.; Dietze, S.; Todorov, K. ClaimsKG: A Knowledge Graph of Fact-Checked Claims. In Proceedings of the 2109 18th International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019; pp. 309–324.
38. Zhen, W.; Jianwen, Z.; Jianlin, F.; Zheng, C. Knowledge Graph Embedding by Translating on Hyperplanes. In Proceedings of the 28th AAAI Conference on Artificial Intelligence, Quebec, Canada, 27–31 July 2014; pp. 1112–1119.
39. Guoliang, J.; Shizhu, H.; Liheng, X.; Kang, L.; Jun, Z. Knowledge Graph Embedding via Dynamic Mapping Matrix. In Proceedings of the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 687–696.
40. Shuai, Z.; Yi, T.; Lina, Y.; Qi, L. Quaternion Knowledge Graph Embeddings. In Proceedings of the 33rd Conference on Neural Information Processing Systems, Vancouver, Canada, 8–14 December 2019.
41. Nickel, M.; Rosasco, L.; Poggio, T. Holographic Embeddings of Knowledge Graphs, arXiv **2015**, arXiv:1510.04935.
42. Zequn, S.; Wei, H.; Qingheng, Z.; Yuzhong, Q. Bootstrapping Entity Alignment with Knowledge Graph Embedding. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 4396–4402.
43. Zhen, W.; Jianwen, Z.; Jianlin, F.; Zheng, C. Knowledge Graph and Text Jointly Embedding. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1591–1601.
44. Xiao, H.; Jingyuan, Z.; Dingcheng, L.; Ping, L.; Knowledge Graph Embedding Based Question Answering. In Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 February 2019; pp. 105–113.
45. Zhu, S.; Jie, Y.; Jie, Z.; Bozzon, A.; Long-Kai, H.; Chi, X. Recurrent Knowledge Graph Embedding for Effective Recommendation. In Proceedings of the 12th ACM Conference on Recommeder System, Vancouver, Canada, 2–7 October 2018; pp. 297–305.
46. Pasquale, M.; Stenetorp, P.; Riedel, S. Convolutional 2D Knowledge Graph Embeddings, arXiv **2017**, arXiv:1707.01476.
47. Bouranoui, Z.; Camacho-Collados, J.; Schockaert, S. Inducing Relational Knowledge from BERT, arXiv **2019**, arXiv:1911.12753.
48. Weijie, L.; Peng, Z.; Zhe, Z.; Zhiruo, W.; Qi, J.; Haotang, D.; Ping, W. K-BERT: Enabling Language Representation with Knowledge Graph, arXiv **2019**, arXiv:1909.07606.
49. Liang, Y.; Chengsheng, M.; Yuan, L. KG-BERT: BERT for Knowledge Graph Completion, arXiv **2019**, arXiv:1909.03193.
50. Eberts, M.; Ulges, A. Span-based Joint Entity and Relation Extraction with Transformer Pre-training, arXiv **2019**, arXiv:1909.07755.
51. Peng, S.; Jimmy, L. Simple BERT Models for Relation Extraction and Semantic Role Labeling, arXiv **2019**, arXiv:1904.05255.
52. Xiaozhi, W.; Tianyu, G.; Zhaocheng, Z.; Zhiyuan, L.; Juanzi, L.; Jian, T. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation, arXiv **2019**, arXiv:1911.06136.
53. Soyeop, Y.; Okran, J. Auto-Growing Knowledge Graph-based Intelligent Chatbot using BERT. ICIC Express Lett. **2020**, 14, 67–73.
54. Ostendorff, M.; Bourgonje, P.; Berger, M.; Moreno-Schneider, J.; Rehm, G.; Gipp, B. Enriching BERT with Knowledge Graph Embeddings for Document Classification, arXiv **2019**, arXiv:1909.08402.

55. Zhibin, L.; Pan, D.; Jian-Yun, N. VGCN-BERT: Augmenting BERT with Graph Embedding for Text Classification. In Proceedings of the 42nd European Conference on Information Retrieval Research, Lisbon, Portugal, 14–17 April 2020; pp. 369–382.
56. English Wikipedia, Open Information Extraction. Available online: https://en.wikipedia.org/wiki/Open_information_extraction (accessed on 15 December 2020).
57. Oren, E. Search needs a shake-up. Nature **2011**, 476, 25–26.
58. Anthony, F.; Luke, Z.; Oren, E. Open question answering over curated and extracted knowledge bases. In The 20th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (KDD '14), New York, NY, USA, 24-27 August 2014; pp. 1156–1165.
59. Stephen, S.; Brendan, R.; Bo, Q.; Shi, X.; Mausam, and Oren, E. Adapting Open Information Extraction to Domain-Specific Relations; AI Magazine: Palo Alto, CA, USA, 2010; Volume 31, pp. 93–102.
60. Michele, B.; Michael, J.C.; Stephen, S.; Matt, B.; Oren, E. Open Information Extraction from the Web. In Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6-12 January 2007; pp. 2670–2676.
61. Anthony, F.; Stephen, S.; Oren, E. Identifying relations for open information extraction. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP '11), Edinburgh, Scotland, UK, 27–31 July 2011; pp. 1535–1545.
62. Gabor, A.; Melvin, J.J.P.; Christopher, D.M. Leveraging linguistic structure for open domain information extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 344–354.
63. Joohong, L.; Sangwoo, S.; Yongsuk, C. Semantic Relation Classification via Bidirectional Networks with Entity-Aware Attention Using Latent Entity Typing. Symmetry **2019**, 11, 785.
64. Gabriel, S.; Julian, M.; Luke, Z.; Ido, D. Supervised Open Information Extraction. In Proceedings of the NAACL-HLT 2018, New Orleans, LA, USA, 1–6 June 2018; pp. 885–895.
65. English Wikipedia, Potal: Coronavirus Disease 2019. Available online: https://en.wikipedia.org/wiki/Portal:Coronavirus_disease_2019 (accessed on 15 October 2020).
66. Kevin, C.; Urvashi, K.; Omer, L.; Christoper, D.M. What Does BERT Look At? An Analysis of BERT's Attention, arXiv **2019**, arXiv:1906.04341.
67. Ganesh, J.; Benoit, S.; Djamé, S. What does BERT Learn About the Structure of Language? In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.
68. Anna, R.; Olga, K.; Anna, R. A Primer in BERTology: What We Know About How BERT Works, arXiv **2020**, arXiv:2002.12327.
69. Olga, K.; Alexey, R.; Anna, R.; Anna, R. Revealing the Dark Secrets of BERT, arXiv **2019**, arXiv:1908.08593.
70. Damina, P.; Gino, B.; Roge, W. Telling BERT's full story: From Local Attention to Global Aggregation, arXiv **2020**, arXiv:2004.05916.
71. Jae-young, J.; Sung-Hyon, M. Roles and Utilization of Attention Heads in Transformer-Based Neural Language Models. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Washington, DC, USA, 5–10 July 2020; pp. 3404–3417.
72. Jesse, V. A Multiscale Visualization of Attention in the Transformer Model, arXiv **2019**, arXiv:1906.05714.
73. Baiyun, C.; Yingming, L.; Ming, C.; Zhongfei, Z. Fine-tune BERT with Sparse Self-Attention Mechanism. 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), Hong Kong, China, 3–7 November 2019; pp. 3548–3553.
74. Qingyun, W.; Manling, L.; Xuan, W.; Nikolaus, P.; Guangxing, H.; Jiawei, M.; Jingxuan, T.; Ying, L.; Haoran, Z.; Weili, L.; et al. COVID-19 Literature Knowledge Graph Construction and Drug Repurosing Report Generation, arXiv **2020**, arXiv:2007.00576.