

# Intent-Controllable Citation Text Generation

Shing-Yun Jung , Ting-Han Lin, Chia-Hung Liao , Shyan-Ming Yuan  and Chuen-Tsai Sun

Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan; thlin.cs08g@nctu.edu.tw (T.-H.L.); aiallen.cs07g@nctu.edu.tw (C.-H.L.); ctsun@cs.nctu.edu.tw (C.-T.S.)  
\* Correspondence: syjung@cs.nycu.edu.tw (S.-Y.J.); smyuan@nycu.edu.tw (S.-M.Y.)

**Abstract:** We study the problem of controllable citation text generation by introducing a new concept to generate citation texts. Citation text generation, as an assistive writing approach, has drawn a number of researchers' attention. However, current research related to citation text generation rarely addresses how to generate the citation texts that satisfy the specified citation intents by the paper's authors, especially at the beginning of paper writing. We propose a controllable citation text generation model that extends a pre-trained sequence to sequence models, namely, BART and T5, by using the citation intent as the control code to generate the citation text, meeting the paper authors' citation intent. Experimental results demonstrate that our model can generate citation texts semantically similar to the reference citation texts and satisfy the given citation intent. Additionally, the results from human evaluation also indicate that incorporating the citation intent may enable the models to generate relevant citation texts almost as scientific paper authors do, even when only a little information from the citing paper is available.

**Keywords:** citation text generation; citation intent; controllable text generation; pre-trained sequence-to-sequence model; natural language processing



**Citation:** Jung, S.-Y.; Lin, T.-H.; Liao, C.-H.; Yuan, S.-M.; Sun, C.-T.

Intent-Controllable Citation Text Generation. *Mathematics* **2022**, *10*, 1763. <https://doi.org/10.3390/math10101763>

Academic Editors: Manuel Vilares-Ferro, Pavel Brazdil and Gaël Dias

Received: 19 April 2022

Accepted: 19 May 2022

Published: 21 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

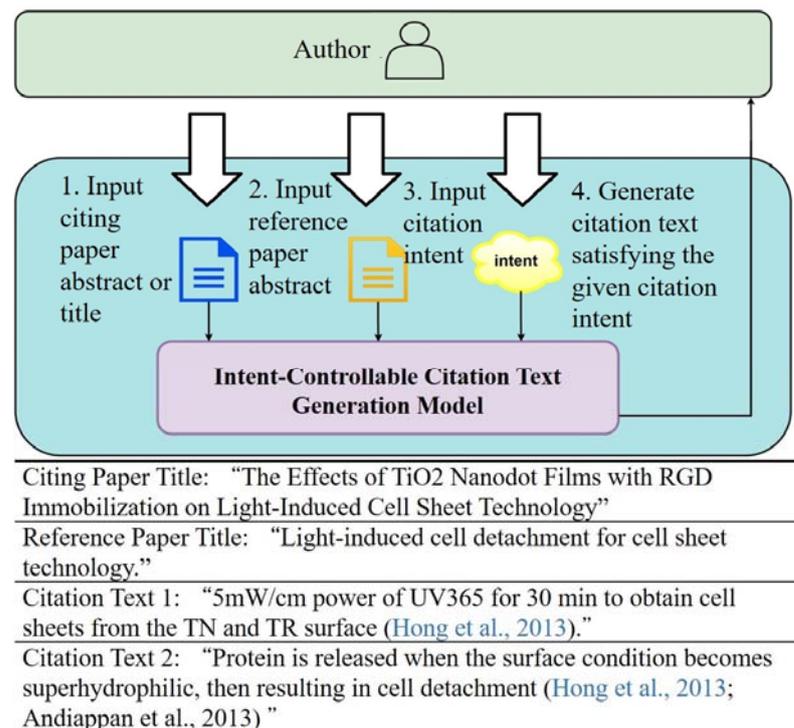
**MSC:** 300093

## 1. Introduction

Citations are crucial in scientific discourse for both readers and authors [1]. For readers, citations not only point out the additional information for the paper's main research problem but also allow readers to verify whether the cited documents support the research argument in the way that the authors claim. For authors, showing their respect for other researchers' ideas by citing their previous work is the authors' responsibility. The aforementioned properties of citation imply that the authors' intent behind the citation text can vary. Boyack et al. [2] analyzed the characteristics of in-text citations in over 5 million articles across five fields of science and reported that the average number of mentions per reference ranged from 1.5 to 1.75. Therefore, one paper can cite another paper a few times in different parts of the full article by using a variety of citation sentences to convey the different relationships between the citing and cited paper, as shown in the lower part of Figure 1. This may be mainly because the authors' citation intent determines the context of the citation text and the information that the citation sentence should convey.

It has been reported that the number of cited references for each scientific paper has increased over the years [2–6]. This soaring number of cited references may imply that researchers should devote more effort to reviewing the relevant literature. Looking through the lens of the cognitive process of academic writing, authors require quite a few mental efforts to complete a citation sentence. In higher educational institutions, the cognitive process and the use of scholarly resources are difficult for students when composing high-quality academic writing [7]. Hence, how to lighten the researchers' burden, especially in paper writing, has become meaningful. Several studies have framed citation text generation as a multisource summarization problem. The source documents require texts from both

the citing paper and the cited paper. Xing et al. [8] proposed a multisource point-generator to generate different citation sentences with the same reference paper using different contexts, namely, a few sentences ahead of and behind the citation sentence from the citing paper. Luu et al. [9] explored how different source documents influence the quality of the generated citation texts using their proposed GPT-2 model [10], which was pre-trained with the subset of S2ORC [11]. However, from the authors' perspective, the challenges of citation text generation still remain. At the beginning of paper writing, the context of the citation text may not be written yet, let alone using the context as the source document for models to generate citation texts. In addition, the approach proposed by Luu et al. [9] can only generate a generic sentence, which describes the relationship between the citing paper and the reference paper. This kind of assistance is likely to be insufficient for researchers who need to mention the same reference paper in different paragraphs.



**Figure 1.** The upper part overviews the intent-controllable citation text generation. The lower part demonstrates that the citation texts for the same reference paper (marked in blue) can be varied depending on the authors' citation intent. We consider the citation intent as a control code to constrain the generated text from sequence-to-sequence models.

Previous inspiring studies related to citation intent classification tasks [12–15] and controllable text generation [10,16–18] have laid a solid foundation for the concept of intent-controllable citation text generation. The former aims to classify the intent of the citation text. The latter aims to generate an output with user-specified constraints. The existing citation intent classification tasks show that essential citation texts can generally be categorized into different citation intents. Since humans can write various citation texts based on their citation intents, it is intuitive to infer that models may generate different citation texts according to the specified citation intents. Hence, we can approach the problem of citation text generation from a more author-centered perspective by leveraging the characteristics of citation intent and controllable text generation.

We argue that the concept of citation intent enables controllable text summarization models to accomplish the task of citation text generation in a more flexible and author-centered way. This is because writing a citation text is a process of comprehending the content of the reference paper, associating the content with the citing paper, and summariz-

ing the content with the specific citation intent. This paper explores the effectiveness and flexibility of intent-controllable citation text generation when the authors only have their paper's abstract or title.

We define intent-controllable citation text generation as follows. Given an abstract or title of a citing paper, an abstract of a reference paper, and the citation intent specified by the author, generate a relevant and fluent citation text satisfying the given citation intent. This definition describes an author who just begins to write a paper. There is no context in the citation texts, and only the paper's abstract or title is available. The cited paper only provides the abstract because the full text is not always available.

The contributions of this work can be summarized as follows:

- We introduce a new concept of intent-controllable citation text generation to generate various citation texts according to authors' citation intent when the same cited paper is given.
- We extend the SciCite dataset for our defined task and the experiments with several models (Both the code and datasets are available at: <https://github.com/BradLin0819/Automatic-Citation-Text-Generation-with-Citation-Intent-Control> (accessed on 23 July 2021)).
- The model performance evaluated by humans and metrics shows that citation intents enhance the effectiveness and flexibility of authors when they write the citation texts at the beginning of paper writing.

The remainder of the paper is structured as follows. In Section 2, we provide an overview of the previous studies related to citation text generation. Section 3 demonstrates how we formulate the problem, how the training dataset is prepared, and what pre-trained language models are used. Section 4 is devoted to the experiment setup and evaluations, including automatic metrics and human evaluation. All the details about implementation are also shown in Section 4. In Section 5, model performances and experimental results are presented, followed by a qualitative comparison to the existing literature.

## 2. Related Work

### 2.1. Generating Texts by Sequence-to-Sequence Models

Sequence-to-sequence (Seq2Seq) models seek to solve the problem of converting a textual sequence  $x = (x_1, \dots, x_T)$  into a textual sequence  $y = (y_1, \dots, y_{T'})$ , for example, machine translation. Encoder-decoder architecture [19] is the foundation for Seq2Seq models. The encoder encodes the input sequence into vectors of hidden states  $z = (z_1, \dots, z_T)$ , which can be considered a summary of the input sequence's information. The decoder then generates the output sequence based on the final hidden state of the encoder. Sutskever et al. [20] used Long-Short-Term-Memory (LSTM) cells [21] to turn the sequence into hidden state vectors. The Seq2Seq architecture problem can be formulated as:

$$p(y_1, \dots, y_{T'} \mid x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t \mid v, y_1, \dots, y_{t-1}) \quad (1)$$

where  $y_t$  is the predicted word at time step  $t$ ,  $v$  is the last hidden state of the encoder,  $T$  is the length of the input sequence, and  $T'$  is the length of the output sequence. The goal of a decoder is to predict the next word given the previous words, which is called an auto-regressive language model.

The Transformer [22], which follows the encoder-decoder architecture, is composed of novel building blocks using the self-attention mechanism to analyze long sequential data. The Transformer has been widely used for natural language processing (NLP), speech recognition tasks, and time-series data modeling because transformer-based models are more parallelizable than LSTM-based models and are better at analyzing longer sequences.

## 2.2. Pre-Trained Sequence-to-Sequence Models in Natural Language Processing

Transfer learning is the concept of pre-training a big neural network on a large dataset with specified unsupervised learning and then fine-tuning the pre-trained model on a subsequent task with a supervised target. Due to the remarkable performance of the Transformer architecture in recent years, numerous Transformer-based pre-trained models have been suggested. These models differ in three significant ways: (1) Transformer modifications, (2) Pre-training objectives, and (3) Pre-training data set. BERT [23] is comprised solely of the encoder component of the Transformer architecture. Pre-training objectives include anticipating the source document's masked tokens and the following phrase. It is trained on a book corpus and the English Wikipedia. GPT-2 [10] contains only the decoder portion of the Transformer and is pre-trained with an auto-regressive language model objective that predicts the next token based on the previous tokens on the WebText corpus. BART [16] and T5 [17] are comprised of the entire Transformer architecture with a few alterations. BART is pre-trained with the purpose of denoising, while T5 is pre-trained with the objective of filling in the blank. It was demonstrated that these two pre-trained models achieve state-of-the-art performance in a variety of text production tasks, including text summarization. In addition, transfer learning can assist limited-data tasks in achieving sufficient performance.

## 2.3. Text Summarization

Citation text generation resembles a variant of text summarization, which summarizes the reference paper and associates it with the citing paper. Researchers proposed automatic text summarizing to acquire knowledge from millions of texts considerably faster. This research attempted to condense the source document into a concise summary. Recent research enhanced the performance of automatic text summarization using neural networks. Existing studies can be divided into two categories: extractive and abstractive summarization.

Extractive summarization works by segmenting the original document into sentences and choosing the best ones to summarize. This method ensures factual coherence between the original document and the summary created. Liu and Lapata [24] addressed extractive summarization as a binary classification problem. They presented BERTSUMEXT, in which the document is encoded using the pre-trained BERT. Then, a classifier is used to predict if a statement should be included in a summary. Zhong et al. [25] formulated extractive summarization as a semantic matching problem. They suggested that the generated summary should be close to the source document in the semantic space.

Abstractive summarization develops new terms rather than using phrases from the source document. The abstractive method is more flexible and human-like. Implementing the abstractive method was difficult in the past. Nonetheless, this obstacle has been steadily overcome by Seq2Seq neural networks. The attentional Seq2Seq model performed well on the abstractive summarization challenge. However, several flaws, such as the inability to generate rare words, which means out of vocabulary (OOV) and word repetition, still remain. See et al. [18] suggested a pointer-generator network that copies infrequent words from the source document to combat OOV. Also, the coverage mechanism was added to keep track of created summaries and reduce recurrence. The pre-trained models have been extended in more and more studies for abstractive text summarization recently.

## 2.4. Citation Text Generation

Xing et al. [8] presented a novel task of generating the citation text given the context of the citing paper and the abstract of the reference paper to reduce the time spent writing the citation text. Moreover, Xing et al. proposed a multisource pointer-generator, which is built upon a pointer-generator [18] but can handle multiple input documents. However, in their approach, the context of the citing paper must be provided. Meanwhile, the citation text has possibly existed. Therefore, their approach lacks the function of automatically generating citation text from scratch. Luu et al. [9] presented a similar task that uses the abstract or introduction of two papers as inputs to generate a citation text. In addition, Luu et al.

proposed a model, extending the GPT-2 pre-trained model with further pre-training on a large-scale scientific paper dataset. Nonetheless, their approach cannot generate multiple citation texts for the same reference. Our work can automatically generate a citation text from scratch given the abstract of two papers and generate multiple citation texts by citation intent involvement.

### 2.5. Controllable Text Generation

Although the neural text generation task has achieved satisfactory results [10,16–18], the existing approaches mainly generate a generic output. However, these approaches have difficulty generating a user-specific output, leading to reduced flexibility of use. Controllable Text Generation aims to generate an output with user-specified constraints. Recently, several studies have incorporated controllable text generation techniques into their tasks. Kobus et al. [26] incorporated domain information into the machine translation task to perform domain-specific translation. Peng et al. [27] proposed a framework to control the ending valence (happy or sad) of the generated stories. Gupta et al. [28] and Wu et al. [29] leveraged controllable text generation techniques to enhance the quality of response in the dialog system. Fan et al. [30] and He et al. [31] performed controllable text summarization by constraining the length, style, or content of the summary.

Nevertheless, controllable text generation techniques have not been incorporated into the citation text generation task. Our work is the first to utilize controllable text generation techniques to generate citation text with author-specified citation intent.

## 3. Methodology

### 3.1. Dataset

To train our citation text generation model, we needed citation texts with human-annotated citation intents. Therefore, we adopted the SciCite dataset [13], which consists of 6627 scientific papers in the computer science and medicine domains. According to Cohan et al. [13], most citation texts can be categorized into three categories: Background, Method, and ResultComparison. The definitions of these three categories are described in Table 1. We needed additional metadata of papers, such as the abstract. The original SciCite dataset does not provide the abstract of the paper. We then used these paper identifiers on Semantic Scholar1 to retrieve the metadata (e.g., abstract, title, etc.) of these papers using Semantic Scholar API (<https://www.semanticscholar.org/product/api>).

**Table 1.** The definitions of citation intent categories.

Intent Category	Definition
Background	“The citation states, mentions, or points to the background information giving more context about a problem, concept approach, topic, or importance of the problem in the field.”
Method	“Making use of a method, tool, approach or dataset”
ResultComparison	“Comparison of the paper’s results/findings with the results/findings of other work”

After retrieving the metadata of papers, each data sample can be represented as (metadata of the citing paper, metadata of the reference paper, citation text, and citation intent label). To build the dataset for our task, we first retained the data splits of the SciCite dataset for training, validation, and testing (8243/916/1861 data samples). Then, we performed the following steps:

1. Removed data samples with duplicate citation texts;
2. To ensure the reliability of our model, we removed data samples with duplicate citing paper and reference paper pairs between the training and validation (test) sets from the training set;

3. We removed data samples with incomplete citation texts because our goal is to generate fluent sentences;
4. Filtered out data samples without the abstract of the reference paper;
5. To normalize the citation markers, we replaced the citation markers that have the current reference paper within them with #REF and other reference papers with #OTHERREF using regular expression operations.

### 3.2. Overview

The basic statistics of each data split are shown in Table 2. In this table, #words denotes the number of words, #sentences denotes the number of sentences, and the standard deviation is in the brackets. These statistics can help us decide the maximum length of input and output sequences to train our models. In addition, most citation texts have only one sentence. This information is helpful in setting the number of selected sentences in extractive methods.

**Table 2.** Basic statistics of each data split.

		Train	Val	Test
Citing Paper	# words (std)	242.58 (101.42)	242.66 (112.14)	250.59 (113.17)
	# sentences (std)	8.99 (3.81)	9.17 (3.95)	9.35 (4.39)
Reference Paper	# words (std)	227.30 (125.71)	226.69 (106.84)	223.99 (114.40)
	# sentences (std)	8.53 (4.78)	8.44 (4.02)	8.39 (4.12)
Citation Text	# words (std)	33.52 (15.35)	33.38 (14.40)	34.60 (15.31)
	# sentences (std)	1.03 (0.17)	1.03 (0.18)	1.03 (0.17)
Intent (%)	Bg/Md/Rs	50.76/34.54/14.51	57.52/29.34/13.12	52.79/33.90/13.30

We also present the distribution (Intent %) of each citation intent category in the bottom row of Table 2, where Bg, Md, and Rs represent Background, Method, and Result-Comparison, respectively. We can observe that data samples of background citation intent account for the majority.

All notations used and their definitions are listed in Table 3 for an easy understanding of the proposed method.

**Table 3.** Summary of notations.

Notation	Definition
CP	Citing paper
RP	Reference paper
$a_{CP}$	Abstract of the citing paper
$t_{CP}$	Title of the citing paper
$a_{RP}$	Abstract of the reference paper
D	Training dataset
C	Predefined citation intent set (control code set)
c	Specified citation intent (control code)
V	Predefined vocabulary
V	Predefined vocabulary size
X	Input sequence of the model
y	Target citation text
$\hat{y}$	Generated citation text

Our approach builds upon controllable text summarization techniques by using the citation intent as a control code, which constrains the style of the generated text. An overview of the proposed method is shown in the upper part of Figure 1. The overall process of our approach can be divided into four steps:

1. The author provides the abstract or title of the paper draft;

2. The author provides the abstract of the paper he/she wants to reference;
3. The author specifies the desired citation intent;
4. Generate the citation text with the given citation intent using our controllable citation text generation model.

### 3.3. Controllable Citation Text Generation Model

We formulate the problem as follows. Given an abstract  $a_{CP}$  or title  $t_{CP}$  of a citing paper CP, an abstract  $a_{RP}$  of a reference paper RP, and a specified citation intent  $c$  from the predefined citation intent set  $C = \{\text{Background, Method, ResultComparison}\}$ , generate the citation text  $\hat{y}$  satisfying the given citation intent. That is, our model should learn to maximize the conditional probability in Equation (2):

$$p(y|a_{CP}, a_{RP}, c; \theta) = \prod_{t=1}^T p(y_t|a_{CP}, a_{RP}, c, y_{<t}; \theta) \tag{2}$$

where  $y$  denotes the target citation text,  $\theta$  denotes the parameters of the model,  $T$  denotes the length of the target citation text,  $y_t$  denotes the token at time step  $t$ , and  $y_{<t} = (y_1, \dots, y_{t-1})$  and  $a_{CP}$  can be replaced by  $t_{CP}$  if only the citing paper’s title is available.

To implement controllable citation text generation and to feed the original input into the pre-trained text generation model, there are four substeps of our pre-processing step.

1. **Concatenation:** While generating citation texts, we take  $a_{CP}/t_{CP}$  and  $a_{RP}$  as input documents. To feed multiple inputs into our model, we concatenated  $a_{CP}/t_{CP}$  and  $a_{RP}$  into a flat sequence  $[a_{CP}/t_{CP}; a_{RP}]$ , where  $[:,:]$  denoted the concatenation.
2. **Prepend control codes:** We assume that the model can discover the association between the input sequence and the output sequence. Accordingly, we employ a simple yet effective technique to enable controllable text generation [26,30–32]. We prepend the textual control codes  $c$  (e.g., citation intent for our case) to the concatenated input  $[a_{CP}/t_{CP}; a_{RP}]$  and form the input sequence  $X$  of the model where  $X = [c; a_{CP}/t_{CP}; a_{RP}]$ . Then, our model is expected to learn to generate the citation text with the desired citation intent. The representation of  $X$  is illustrated in Figure 2.
3. **Tokenization:** Because the model can only input numerical values, we cannot directly feed the **textual** input into the model. Therefore, we need to tokenize the input sequence  $X$  into tokens  $(x_1, x_2, \dots, x_m)$ , and output sequence  $y$  into tokens  $(y_1, y_2, \dots, y_n)$  where  $m, n$  is the length of the sequence. Then, we convert these tokens into indices using predefined vocabulary  $V$ , which consists of the most common tokens and the mapping between the token and the index.
4. **Pad or Truncate:** To have the same input and output sequence size for the model, we fix the input and output sequence length. We set the maximal length of the input sequence to 1024 and the maximal length of the output sequence to 128. First, we add  $\langle s \rangle$ , one special token, to the start of the sentence and  $\langle /s \rangle$ , the other special token, to the end of the sentence to make the model identify the start and the end of the sentence. Then, the sequence that is longer than the maximal length will be truncated. Furthermore, we pad sequences in the same batch with the special token  $\langle \text{pad} \rangle$  to the longest sequence in the batch.



Figure 2. Illustration of input representation.

### 3.3.1. Pre-trained Text Generation Model

We choose two Transformer-based pre-trained sequence-to-sequence models, BART [16] and T5 [17], which have been shown to achieve state-of-the-art results on several NLP tasks. The size of the citation intent annotated dataset is appropriate for fine-tuning the two pre-trained sequence-to-sequence models.

We utilize these two Transformer-based pre-trained models to take the pre-processed sequence  $X$  (Figure 2) as input and generate the citation text  $\hat{y}$ . First, the input sequence  $X$  is tokenized and converted to indices. The input embedding block converts these indices into dense vectors, and positional embeddings are added to keep the information in the order of the sequence. Then, the Transformer encoder layers encode the input sequence to vectors of hidden states. We feed the vectors of hidden states of the input sequence and the previous output token ( $\langle s \rangle$  is fed when no previous output token exists) into the Transformer decoder layers. While predicting the next token, the decoder attends to the hidden states of the input sequence and hidden states of the already generated output token. However, the output of decoder layers are vectors of floats of dimension  $d$ , and they should be turned into real tokens. Thus, the Linear block projects the output of the Transformer decoder into vectors of the dimension of vocabulary size  $|V|$ , and the softmax function is used to output the probability of the token  $\hat{y}_t$  at time step  $t$ :

$$p(\hat{y}_t) = \text{softmax}(W_o o_t) \quad (3)$$

where  $p(\hat{y}_t) \in \mathbb{R}^{|V|}$  is the probability distribution over vocabulary,  $W_o \in \mathbb{R}^{|V| \times d}$  is a trained projection matrix (Linear block), and  $o_t$  is the output from the hidden states of the decoder layers. We can obtain the token  $\hat{y}_t$  by taking the index of the maximal probability of  $p(\hat{y}_t)$ . This decoding method is called greedy decoding. We applied this decoding method during the training phase.

### 3.3.2. Model Training

Given a training set  $D$  of (preprocessed input sequence, target citation text) pairs  $\{(X^{(1)}, y^{(1)}), \dots, (X^{(i)}, y^{(i)})\}$ , we train our model by minimizing negative-log-likelihood (NLL) loss:

$$L_{NLL} = - \sum_{i=1}^{|D|} \sum_{t=1}^T \log p(y_t^{(i)} | X^{(i)}, y_{<t}^{(i)}; \theta) \quad (4)$$

where  $L_{NLL}$  denotes the NLL loss,  $D$  denotes the training set,  $|D|$  denotes the number of training samples,  $T$  denotes the length of the target citation text,  $y_t^{(i)}$  denotes the token at time step  $t$  of the  $i$ -th training sample,  $X^{(i)}$  denotes the input sequence of the  $i$ -th training sample,  $y_{<t}^{(i)}$  denotes the token generated before time step  $t$  of the  $i$ -th training sample, and  $\theta$  denotes the parameters of the citation text generation model.

### 3.3.3. Model Inference

When we want to cite the reference paper with the specific citation intent, we can prepend the corresponding citation intent control codes to the concatenated sequence of the abstract of the author's draft paper and the abstract of the reference paper. Then, we feed this sequence into the well-trained model, and the model outputs a citation text with the desired citation intent. While generating an output sequence, the model predicts a sequence of probability vectors of dimension  $|V|$ .

The simplest method to decide the output token at each step is selecting the index of maximal probability. Nonetheless, this method can only generate a locally optimal sequence. To generate an approximately globally optimal sequence, we apply the beam search decoding method.

Finally, we select a sequence of tokens with the maximal sum of log probability as the generated citation text. While generating sequences using the beam search algorithm, we have to keep track of  $k$  sequences with the maximal length of  $L$ , and at each time step  $t$ ,

we need to search the vocabulary space of dimension  $|V|$ . Thus, the time complexity of the beam search algorithm is  $O(k \cdot L \cdot |V|)$ . The flowchart of the whole citation sentence generation process is illustrated in Figure 3.

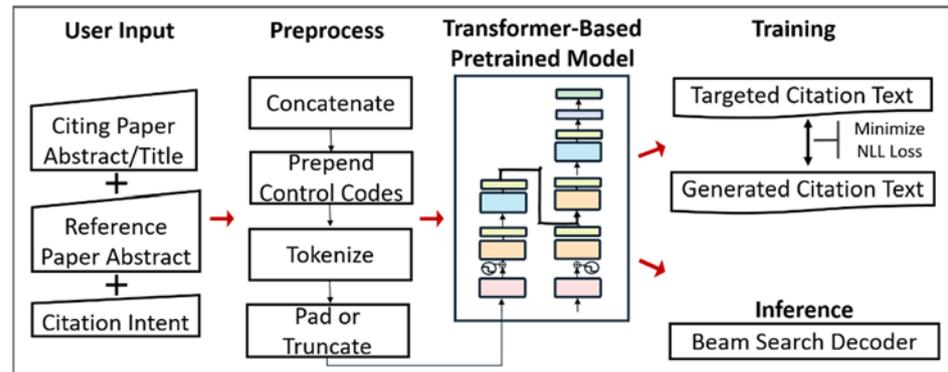


Figure 3. Flowchart of the whole citation sentence generation process.

#### 4. Experiments

We conducted the following experiments to verify the effectiveness and flexibility of using citation intents. T5 and BART are the two based models we chose. Three different tuples of training samples were used for fine-tuning, including 1.  $(a_{CP}, a_{RP})$ , 2.  $(c, a_{CP}, a_{RP})$ , 3.  $(c, t_{CP}, a_{RP})$ . In total, six models are to be fine-tuned. They are BART-no\_intent, BART- $a_{CP} + a_{RP}$ , BART- $t_{CP} + a_{RP}$ , T5-no\_intent, T5- $a_{CP} + a_{RP}$ , and T5- $t_{CP} + a_{RP}$ . BART-no\_intent and T5-no\_intent are two baseline models to be compared with other models fine-tuned with citation intents. These two models can generate only one citation sentence given the abstracts of the citing and reference paper during inference. The SciCite dataset allows only one citation intent to map only one citation text. The other four models can generate three citation sentences based on the specified citation intent and the given abstracts or title of the citing and the abstract of the reference paper during inference.

##### 4.1. Evaluation Metrics

The effectiveness of citation intent can be evaluated by comparing the performance of the two models. One model was fine-tuned using citation intents and the other using no citation intents. The performance was measured by the similarity between generated citation texts and the original reference citation texts in the citing paper.

**ROUGE:** We used  $F_1$  scores for ROUGE-1/2/L [33], which are commonly used in the text summarization task.

**EXT-Oracle** is the upper-bound extractive summarization method that was also used as a baseline for model performance comparison [24].

**SciBERTScore:** To compensate for the shortcoming of ROUGE, we further calculated the  $F_1$  score for SciBERTScore by replacing BERT's contextualized embedding in BERTScore [34] with SciBERT's [35] configuration.

Additionally, we evaluated the flexibility by verifying whether models with the intent control can generate citation texts satisfying the specified citation intent.

**Citation Intent Accuracy** was used to measure the citation intent correctness of generated citation texts. We used the citation intent classifier provided by Cohan et al. [13] to label the citation intent of generated citation texts because this classifier has been proven to achieve an 84 macro- $F_1$  score on the SciCite dataset. The calculation of Citation Intent Accuracy is as follows:

$$\text{Citation Intent Accuracy} = \frac{\sum_{i=1}^N I(\text{Pred}(\hat{y}^{(i)}) = c^{(i)})}{N}$$

where

$$I(.) = \begin{cases} 1, & \text{if the condition satisfies} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$N$  is the number of test samples,  $\hat{y}^{(i)}$  denotes the generated citation text of the  $i$ -th test sample,  $c$  denotes the specified citation intent of the  $i$ -th test sample,  $\text{Pred}(\cdot)$  denotes the predicted citation intent by the citation intent classifier.

#### 4.2. Implementation Detail

The development environment is as follows: OS: Ubuntu 20.04, OS memory 64 GB, GPU: RTX2090, GPU memory 24 GB, python3.7, CUDA 11.0. All the models were implemented using the Hugging Face's Transformers library [36] and PyTorch 1.9.0. We used the BART-base configuration and T5-base configuration for all BART-based models and T5-based models, respectively. We used the Adam [37] optimizer with a learning rate of  $1 \times 10^{-5}$ ,  $\beta_1$  of 0.9,  $\beta_2$  of 0.999,  $\epsilon$  of  $1 \times 10^{-8}$ , batch size of 4 for BART-based models, batch size of 2 for T5-based models, gradient accumulation steps of 4, and gradient clipping norm of 1.0 to fine-tune all 6 models. Each model was trained until the maximum of 40,000 steps with an early stopping strategy of patience 3. The best model for each tuple of training samples was selected based on the validation loss. We set the beam size  $k$  to 4, which was tuned based on the validation set ROUGE-1.

#### 4.3. Human Evaluation

We derived our human evaluation from Luu et al.'s work [9]. The quality of the generated citation texts was measured by how correct, specific, and plausible these texts were. We continued to use the definitions of correct, specific, and plausible defined by Luu et al. The questions for the evaluators are below:

- (Correct) Does the citation sentence correctly express the factual relationship between the citing paper and the cited paper?
- (Specific) Does the citation sentence describe a specific relationship between the two papers, or is it vague enough to be used to cite many other papers?
- (Plausible) Does the citation sentence have the same topicality and tone in writing as the citing paper?

Evaluators could answer "Yes (Specific)", "No (Vague)", or "Unsure". The answer with "Unsure" is invalid. The scores were calculated by the proportion of "Yes" to all valid answers. We also calculated the scores for correct and specific (Corr&Spec), which means that both the answers to correct and specific require "Yes".

In addition, whether the generated citation texts satisfy the specified citation intent was also evaluated by asking the evaluator the following question.

- (Intent) In which part of the paper does the citation sentence belong, introduction, method, or result?

The scores of the citation intent were calculated in a similar way to the citation intent accuracy. If the evaluator's answer equals the given citation intent, the answer receives one point. Otherwise, the answer receives zero. The final scores are the sum of the total points divided by the total number of valid answers.

We randomly selected 32 samples of computer science fields from 1861 testing data samples. The citation texts written by human authors are set as the gold standard samples. The gold standard samples were also evaluated to assess the quality of evaluation. We recruited 30 graduate students from the computer science department for evaluation. Half of the evaluators had experience in writing scientific papers. The agreement of evaluation was calculated by Krippendorff's alpha [38], which is denoted as "Agr." in the results of human evaluation.

## 5. Results and Discussion

The results of the evaluation by metrics are shown in Table 4. As shown in Table 4, models incorporating the citation intent achieve better performance for all metrics. These results demonstrate the effectiveness of the citation intent. It is worth noting that BART-a<sub>CP</sub> + a<sub>RP</sub> and BART-t<sub>CP</sub> + a<sub>RP</sub> achieve over 63.8 SciBERTScore, which is close to the model performance of the similar task, scientific paper summarization [39]. Additionally, most fine-tuned models slightly outperform the upper-bound extractive method, EXT-Oracle, on all metrics except for ROUGE-2. For the higher ROUGE-2 score of EXT-Oracle, we conjecture that many proper nouns are composed of two words. Abstractive methods tend to generate novel words, which might lead to the poorer results of ROUGE-2.

**Table 4.** The evaluation results by metrics. R-1 denotes ROUGE-1, R-2 denotes ROUGE-2, R-L denotes ROUGE-L, SciBS denotes SciBERTScore, and Intent Acc denotes citation intent accuracy. The bold numbers show the highest scores among all fine-tuned models.

	R-1	R-2	R-L	SciBS	Intent Acc.
EXT-Oracle	22.21	<b>4.96</b>	16.06	62.35	
BART-no_intent	22.74	3.03	15.85	63.16	
BART-a <sub>CP</sub> + a <sub>RP</sub>	23.64	3.45	16.96	63.81	<b>88.34</b>
BART-t <sub>CP</sub> + a <sub>RP</sub>	<b>24.05</b>	3.55	<b>16.98</b>	<b>63.82</b>	87.91
T5-no_intent	21.84	2.63	15.15	61.78	
T5-a <sub>CP</sub> + a <sub>RP</sub>	22.97	3.04	15.90	62.61	77.52
T5-t <sub>CP</sub> + a <sub>RP</sub>	23.85	3.44	16.59	63.37	82.99

In terms of citation intent accuracy, both BART-based and T5-based models achieve over 77% citation intent accuracy. These results indicate that the model can generate citation texts that satisfy the given citation intents. These results also imply that the flexibility of the model is increased by incorporating citation intents as control codes.

Table 5 lists the results of evaluation by metrics for each intent. As shown in Table 5, our models have better performance in generating ResultComparison citation texts. We conjecture that the purpose of writing ResultComparison citation texts is more specific and only compares the experimental results between two papers. However, the Background and Method citation intent can be divided into more fine-grained intents. For example, the citation text with Background citation intent can describe the research problem's motivation or the definition of the keywords in previous papers.

**Table 5.** The evaluation results by metrics for each intent. The bold numbers show the highest scores among all fine-tuned models for each intent.

Background	R-1	R-2	R-L	SciBS
BART-a <sub>CP</sub> + a <sub>RP</sub>	22.99	3.24	<b>16.93</b>	<b>64.20</b>
BART-t <sub>CP</sub> + a <sub>RP</sub>	<b>23.38</b>	<b>3.30</b>	16.83	64.13
T5-a <sub>CP</sub> + a <sub>RP</sub>	22.13	2.65	15.62	62.73
T5-t <sub>CP</sub> + a <sub>RP</sub>	23.10	3.08	16.32	63.49
Method	R-1	R-2	R-L	SciBS
BART-a <sub>CP</sub> + a <sub>RP</sub>	22.85	2.90	<b>16.02</b>	<b>62.12</b>
BART-t <sub>CP</sub> + a <sub>RP</sub>	22.90	3.16	15.90	62.09
T5-a <sub>CP</sub> + a <sub>RP</sub>	22.30	2.97	15.21	61.09
T5-t <sub>CP</sub> + a <sub>RP</sub>	<b>23.20</b>	<b>3.34</b>	15.79	61.75
ResultCompare	R-1	R-2	R-L	SciBS
BART-a <sub>CP</sub> + a <sub>RP</sub>	29.46	<b>5.78</b>	<b>20.46</b>	<b>67.26</b>
BART-t <sub>CP</sub> + a <sub>RP</sub>	<b>29.57</b>	5.63	20.42	67.13
T5-a <sub>CP</sub> + a <sub>RP</sub>	27.87	4.73	18.70	65.97
T5-t <sub>CP</sub> + a <sub>RP</sub>	28.51	5.04	19.62	66.99

According to Table 6, both BART-based and T5-based models obtain higher quality than the gold standard regarding plausibility and intent. T5- $t_{CP} + a_{RP}$  receives higher quality than the gold standard regarding specific and marginally lower quality than the gold standard regarding correct and both correct and specific

**Table 6.** Results of human evaluation. The bold numbers show the highest scores among all fine-tuned models except the gold standard.

	Plaus.	Spec.	Int.	Corr.	Cor&Spec.	Agr.
B- $a_{CP} + a_{RP}$	0.93	0.78	0.65	0.70	0.46	0.50
B- $t_{CP} + a_{RP}$	0.84	0.62	<b>0.74</b>	0.73	0.50	0.74
T5- $a_{CP} + a_{RP}$	<b>0.98</b>	0.68	0.60	0.70	0.51	0.61
T5- $t_{CP} + a_{RP}$	0.84	<b>0.80</b>	0.69	<b>0.78</b>	<b>0.65</b>	0.71
Gold	0.92	0.79	0.67	0.81	0.69	0.69
Agr.	0.69	0.71	0.71	0.77	0.71	

Both evaluations of humans and metrics show that most of the modes fine-tuned with tuples of  $(c, t_{CP}, a_{RP})$  outperform the other models fine-tuned in the other way. We conjecture that while writing citation texts with background and method citation intent, the content of the reference paper is much more important because it aims to elaborate on the state of current research or use the method proposed by the reference paper. Therefore, the content of the citing paper has a relatively low influence. However, when writing citation texts with ResultComparison citation intent, the author needs to know the experimental results of citing paper and reference paper. Thus, the abstract of the citing paper is more important for generating citation text with ResultComparison citation intent.

Evaluation metrics and human evaluation measure the quality of generated citation texts from different perspectives. Higher scores from evaluation metrics imply that the model may generate some citation texts that might have a more similar concept to the referenced citation text written by a human. Higher scores from human evaluation suggest that more evaluators sense the attributes of citation texts when they read the citation texts generated by models.

In Table 6, two possible explanations for the lower intent accuracy of the gold standard may be (1). The intents as prepend control codes guide the model to learn the most general sentence structures or writing styles for the corresponding intents. These sentence structures and writing styles signify the authors' intents for readers. (2). Human-written citation texts may have diverse ways of expressing ideas. The citation intent is not necessarily explicit in the sentence structures or writing styles. Thus, it is harder for evaluators to correctly classify the citation intent of the human-written citation sentences.

This study is compared to two earlier studies on citation text generation in Table 7. Attributed to the reason that each of the three studies addresses a specific issue, their training data pairs and input contents are unique. Xing et al. attempted to produce more data points for the automatic citation texts generation task. Luu et al. focused on the generation of the sentences to explain the relationship between two papers. They used citation sentences as proxies for the explanations. Our work aims at lightening the authors' burden of paper writing by leveraging the concept of citation text generation with intent control. Despite the fact that the R-L score is utilized by all three research, the various training data sets limit the quantitative comparison between the three R-L scores. The method proposed by Xing et al. generates citation texts with the highest R-L scores. One apparent reason is that they used three sentences preceding and three sentences following the citation text as input to the context of the citation sentence. However, the process of writing is sequential. Generally, the three sentences following the citation sentence are written after the citation sentence has been completed. Therefore, the input contents may be ineffective in terms of supporting authors with the paper-writing process.

**Table 7.** Comparison with previous studies.

	Input	Training Dataset	Best Model	Metrics
Xing et al. [8]	<ul style="list-style-type: none"> <li>• Citing paper’s context</li> <li>• Three sentences before and three sentences after the citation text</li> <li>• Cited paper’s abstract</li> </ul>	ACL Anthology Network corpus <ul style="list-style-type: none"> <li>• 16.675 K papers</li> <li>• 85 K citation sentences</li> </ul>	Multi-Source Pointer-Generator Network with Cross Attention	R-1: 26.28 R-2: 7.05 R-L: 20.49
Luu et al. [9]	<ul style="list-style-type: none"> <li>• Citing paper’s introduction</li> <li>• The sorted words with 100 highest tf-idf score in cited paper</li> </ul>	S2-GORC <ul style="list-style-type: none"> <li>• 154 K papers</li> <li>• 622 K citation sentences</li> </ul>	Pritrained-GPT-2	R-L-2: 12.3
This work	<ul style="list-style-type: none"> <li>• Citing paper’s title or abstract</li> <li>• Cited paper’s abstract</li> <li>• Citation intent</li> </ul>	Extended SciCite <ul style="list-style-type: none"> <li>• 6.627 K papers</li> <li>• 11 K citation sentences</li> </ul>	BART- $t_{CP}$ + $a_{RP}$	R-1: 24.05 R-2: 3.55 R-L: 16.98

Regarding the efforts in model training, our model requires the least training efforts among the 3 studies by using only 11 K citation sentences from 6.627 K papers to fine-tune transformer-based pre-trained language models. By contrast, both Xing et al. and Luu et al. trained their models from scratch and used approximately 7 to 57 times the quantity of citation sentences we used to train our models. Lower training efforts indicate a broader range of applications because authors from various fields can tailor their citation text generation models by fine-tuning the language models using a certain amount of papers in their fields.

The training dataset used in this study limits, however, certain situations in real writing scenarios. Considering real writing scenarios, authors may have more fine-grained intents than the three coarse intents we use in this work, and authors may know precisely which paragraph or lines in the reference paper to cite. The input contents related to reference papers can be other paragraphs and should not necessarily be the abstract.

We list some citation texts generated by our best model in Table 8. For the generated Background citation text, our model captures the main finding of the reference paper and takes it as the citation text. For the generated Method citation text, our model recognizes the experimental method used in the reference paper and finds how this method can be applied to the citing paper. Moreover, for the generated ResultComparison, our model compares the findings of the two papers and generates the citation text. However, the generated ResultComparison citation text is slightly unfluent due to the redundant words “in 3D space”. More examples of generated citation texts are shown in Supplementary Materials [40–57].

**Table 8.** Examples of generated citation texts.

**Citing paper title:** “Exogenous orienting of crossmodal attention in 3-D space: Support for a depth-aware crossmodal attentional system” [58]

**Citing paper abstract:** “The aim of the present study was to investigate exogenous crossmodal orienting of attention in three-dimensional (3-D) space. Most studies in which the orienting of attention has been examined in 3-D space concerned either exogenous intramodal or endogenous crossmodal attention. Evidence for exogenous crossmodal orienting of attention in depth is lacking. Endogenous and exogenous attention are behaviorally different, suggesting that they are two different mechanisms. We used the orthogonal spatial-cueing paradigm and presented auditory exogenous cues at one of four possible locations in near or far space before the onset of a visual target. Cues could be presented at the same (valid) or at a different (invalid) depth from the target (radial validity), and on the same (valid) or on a different (invalid) side (horizontal validity), whereas we blocked the depth at which visual targets were presented. Next to an overall validity effect (valid RTs < invalid RTs) in horizontal space, we observed an interaction between the horizontal and radial validity of the cue: The horizontal validity effect was present only when the cue and the target were presented at the same depth. No horizontal validity effect was observed when the cue and the target were presented at different depths. These results suggest that exogenous crossmodal attention is depth-aware, and they are discussed in the context of the supramodal hypothesis of attention.”

Table 8. Cont.

<b>Reference paper title:</b> “Spatial cuing in a stereoscopic display: Evidence for a depth-aware attentional focus.” [59]
<b>Reference paper abstract:</b> “Two experiments were conducted to explore whether attentional selection occurs in depth, or whether attentional focus is depth blind, as suggested by Ghiradelli and Folk (1996). In Experiment 1, observers viewed stereoscopic displays in which one of four spatial locations was cued. Two of the locations were at a near-depth location and two were at a far-depth location, and a single target was presented along with three distractors. The results indicated a larger cost in reaction time for switching attention in x,y and depth than in x,y alone, supporting a depth-aware attentional spotlight. In Experiment 2, no distractors were present, similar to the displays used by Ghiradelli and Folk. In this experiment, no effect for switching attention in depth was found, indicating that the selectivity of attention in depth depends on the perceptual load imposed on observers by the tasks and displays.”
<b>Original Citation Text:</b> “Our findings are in line with the results from studies in which exogenous intramodal orienting of attention was investigated in 3-D space #REF.”
<b>Generated Citation Texts</b>
<b>Intent: Background</b> “It has been shown that attentional selection in depth depends on the perceptual load imposed on observers by the tasks and displays #REF.”
<b>Intent: Method</b> “We used the orthogonal spatial cueing paradigm #REF to investigate the spatial-cueing of attention in 3-D space.”
<b>Intent: ResultComparison</b> “Our findings are consistent with previous studies in which the orienting of attention in 3D space has been examined in 3-D space #REF.”

## 6. Conclusions and Future Work

This paper introduces the concept of intent-controllable citation text generation to increase the effectiveness and flexibility of citation text generation. Not only does the introduced concept give more author-centered paper-writing assistance than existing studies, but it also provides a handy way to customize citation text generation models for various fields by leveraging the characteristic of citation intents and pre-trained sequence-to-sequence models. The experimental results demonstrate that by incorporating citation intent, the fine-tuned models can generate relevant citation texts that satisfy the specified citation intents, even when only a little information from the citing paper is available.

The flexibility of our method may be expanded by collecting data with more categories of citation intent. Our approach may generate more diverse citation texts based on the data with more citation intent categories. Moreover, to mitigate this limitation, we will also investigate other attributes of the citation text with different citation intents in future work and use them as control codes. We also believe that if the authors provide the most relevant paragraph of the reference paper that they want to cite, it can help the model generate more accurate citation texts, satisfying the authors’ needs.

Currently, there is no large-scale dataset composed of citation text and its corresponding most relevant paragraph data pairs to train the model due to the highly professional annotation process. Therefore, the question of how to efficiently build such a dataset is also a challenging problem. To help the author use the automatic citation text generation function more conveniently, in future work, we will build a complete web service or integrate this function into existing writing tools.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math10101763/s1>, Table S1: Example of generated citation texts.

**Author Contributions:** Conceptualization, T.-H.L.; methodology, T.-H.L. and C.-H.L.; validation, S.-Y.J.; formal analysis, S.-Y.J.; investigation, S.-Y.J.; writing—original draft preparation, S.-Y.J.; writing—review and editing, S.-M.Y. and C.-T.S.; visualization, S.-Y.J. and C.-H.L.; supervision, S.-M.Y. and C.-T.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology of Taiwan (grant no. 108-2511-H-009-009-MY3).

**Data Availability Statement:** Both the code and datasets are available at: <https://github.com/BradLin0819/Automatic-Citation-Text-Generation-with-Citation-Intent-Control> (accessed on 23 July 2021).

**Acknowledgments:** The authors of this study sincerely thank Cheng-Han Gao for setting the human evaluation process and reviewing the collected data.

**Conflicts of Interest:** The authors declare no competing interest.

## References

1. Santini, A. The importance of referencing. *J. Crit. Care Med.* **2018**, *4*, 3–4. [[CrossRef](#)]
2. Boyack, K.W.; van Eck, N.J.; Colavizza, G.; Waltman, L. Characterizing in-text citations in scientific articles: A large-Scale analysis. *J. Informetr.* **2018**, *12*, 59–73. [[CrossRef](#)]
3. Bornmann, L.; Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2215–2222. [[CrossRef](#)]
4. Hsiao, T.M.; Chen, K.-H. How authors cite references? A study of characteristics of in-text citations. *Proc. Assoc. Inf. Sci. Technol.* **2018**, *55*, 179–187. [[CrossRef](#)]
5. Nicolaisen, J.; Frandsen, T.F. Number of references: A large-scale study of interval ratios. *Scientometrics* **2021**, *126*, 259–285. [[CrossRef](#)]
6. Ucar, I.; López-Fernandino, F.; Rodriguez-Ulibarri, P.; Sesma-Sanchez, L.; Urrea-Micó, V.; Sevilla, J. Growth in the number of references in engineering journal papers during the 1972–2013 period. *Scientometrics* **2014**, *98*, 1855–1864. [[CrossRef](#)]
7. Akin, I.; MurrellJones, M. Closing the Gap in Academic Writing Using the Cognitive Load Theory. *Lit. Inf. Comput. Educ. J.* **2018**, *9*, 2833–2841. [[CrossRef](#)]
8. Xing, X.; Fan, X.; Wan, X. Automatic generation of citation texts in scholarly papers: A pilot study. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 6181–6190.
9. Luu, K.; Wu, X.; Koncel-Kedziorski, R.; Lo, K.; Cachola, I.; Smith, N.A. Explaining Relationships Between Scientific Documents. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Bangkok, Thailand, 2–4 August 2021; Volume 1, pp. 2130–2144.
10. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
11. Lo, K.; Wang, L.L.; Neumann, M.; Kinney, R.; Weld, D.S. S2ORC: The Semantic Scholar Open Research Corpus. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 4969–4983.
12. Abu-Jbara, A.; Ezra, J.; Radev, D. Purpose and polarity of citation: Towards nlp-based bibliometrics. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; pp. 596–606.
13. Cohan, A.; Ammar, W.; van Zuylen, M.; Cady, F. Structural Scaffolds for Citation Intent Classification in Scientific Publications. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 3586–3596.
14. Jha, R.; Abu-Jbara, A.; Qazvinian, V.; Radev, D.R. NLP-driven citation analysis for scientometrics. *Nat. Lang. Eng.* **2017**, *23*, 93–130. [[CrossRef](#)]
15. Valenzuela, M.; Ha, V.; Etzioni, O. Identifying meaningful citations. In Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–26 January 2015.
16. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Seattle, WA, USA, 5–10 July 2020; pp. 7871–7880.
17. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 1–67.
18. See, A.; Liu, P.J.; Manning, C.D. Get to the point: Summarization with pointer-generator networks. *arXiv* **2017**, arXiv:1704.04368.
19. sooftware/seq2seq. Available online: <https://github.com/sooftware/seq2seq> (accessed on 21 January 2022).
20. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3104–3112.
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
22. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
23. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
24. Liu, Y.; Lapata, M. Text Summarization with Pretrained Encoders. 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. 3730–3740.
25. Zhong, M.; Liu, P.; Chen, Y.; Wang, D.; Qiu, X.; Huang, X. Extractive summarization as text matching. *arXiv* **2020**, arXiv:2004.08795.

26. Kobus, C.; Crego, J.M.; Senellart, J. Domain Control for Neural Machine Translation. In Proceedings of the International Conference Recent Advances in Natural Language Processing, Hissar, Bulgaria, 9–11 September 2013; pp. 372–378.
27. Peng, N.; Ghazvininejad, M.; May, J.; Knight, K. Towards controllable story generation. In Proceedings of the First Workshop on Storytelling, New Orleans, LA, USA, 5–6 June 2018; pp. 43–49.
28. Gupta, P.; Bigham, J.P.; Tsvetkov, Y.; Pavel, A. Controlling dialogue generation with semantic exemplars. *arXiv* **2020**, arXiv:2008.09075.
29. Wu, Z.; Galley, M.; Brockett, C.; Zhang, Y.; Gao, X.; Quirk, C.; Koncel-Kedziorski, R.; Gao, J.; Hajishirzi, H.; Ostendorf, M. A controllable model of grounded response generation. *arXiv* **2020**, arXiv:2005.00613.
30. Fan, A.; Grangier, D.; Auli, M. Controllable Abstractive Summarization. In Proceedings of the 2nd Workshop on Neural Machine Translation and Generation, Melbourne, Australia, 20 July 2018; pp. 45–54.
31. He, J.; Kryściński, W.; McCann, B.; Rajani, N.; Xiong, C. Ctrlsum: Towards generic controllable text summarization. *arXiv* **2020**, arXiv:2012.04281.
32. Tan, B.; Qin, L.; Xing, E.; Hu, Z. Summarizing Text on Any Aspects: A Knowledge-Informed Weakly-Supervised Approach. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Virtual Event, 16–20 November 2020; pp. 6301–6309.
33. Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Text Summarization Branches Out, Barcelona, Spain, 25–26 July 2004; pp. 74–81.
34. Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K.Q.; Artzi, Y. BERTScore: Evaluating Text Generation with BERT. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
35. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3615–3620.
36. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 38–45.
37. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
38. Hayes, A.F.; Krippendorff, K. Answering the Call for a Standard Reliability Measure for Coding Data. *Commun. Methods Meas.* **2007**, *1*, 77–89. [[CrossRef](#)]
39. Gabriel, S.; Bosselut, A.; Da, J.; Holtzman, A.; Buys, J.; Lo, K.; Celikyilmaz, A.; Choi, Y. Discourse Understanding and Factual Consistency in Abstractive Summarization. *arXiv* **2019**, arXiv:1907.01272.
40. Zhang, J.; Tang, J.; Liu, L.; Li, J. A mixture model for expert finding. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Osaka, Japan, 20–23 May 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 466–478.
41. Buckley, C.; Voorhees, E.M. Retrieval evaluation with incomplete information. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, 25–29 July 2004; pp. 25–32.
42. Woodard, J.L.; Seidenberg, M.; Nielson, K.A.; Miller, S.K.; Franczak, M.; Antuono, P.; Douville, K.L.; Rao, S.M. Temporally graded activation of neocortical regions in response to memories of different ages. *J. Cogn. Neurosci.* **2007**, *19*, 1113–1124. [[CrossRef](#)] [[PubMed](#)]
43. Alwagait, E.; Shahzad, B.; Alim, S. Impact of social media usage on students academic performance in Saudi Arabia. *Comput. Hum. Behav.* **2015**, *51*, 1092–1097. [[CrossRef](#)]
44. Ananthakrishnan, S.; Narayanan, S. Improved speech recognition using acoustic and lexical correlates of pitch accent in a n-best rescoring framework. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Honolulu, HI, USA, 15–20 April 2007; pp. IV-873–IV-876.
45. Bošnjak, M.; Rocktäschel, T.; Naradowsky, J.; Riedel, S. Programming with a differentiable forth interpreter. *Proc. Mach. Learn. Res.* **2017**, *70*, 547–556.
46. Bouachir, W.; Kardouchi, M.; Belacel, N. Fuzzy indexing for bag of features scene categorization. In Proceedings of the 2010 5th International Symposium On I/V Communications and Mobile Network, Rabat, Morocco, 30 September–2 October 2010; pp. 1–4.
47. Bouker, M.A.; Hervet, E. Retrieval of images using mean-shift and gaussian mixtures based on weighted color histograms. In Proceedings of the 2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems, Dijon, France, 28 November–1 December 2011; pp. 218–222.
48. Chiang, C.-Y.; Siniscalchi, S.M.; Chen, S.-H.; Lee, C.-H. Knowledge integration for improving performance in LVCSR. In Proceedings of the INTERSPEECH, 2013 (14th Annual Conference of the International Speech Communication Association), Lyon, France, 25–29 August 2013; pp. 1786–1790.
49. Dawelbait, G.; Mezher, T.; Woon, W.L.; Henschel, A. Taxonomy based trend discovery of renewable energy technologies in desalination and power generation. In Proceedings of the PICMET 2010 Technology Management for Global Economic Growth, Phuket, Thailand, 18–22 July 2010; pp. 1–8.
50. Doleck, T.; Lajoie, S. Social networking and academic performance: A review. *Educ. Inf. Technol.* **2018**, *23*, 435–465. [[CrossRef](#)]
51. Doleisch, H. SimVis: Interactive visual analysis of large and time-dependent 3D simulation data. In Proceedings of the 2007 Winter Simulation Conference, Washington, DC, USA, 9–12 December 2007; pp. 712–720.

52. Henschel, A.; Casagrande, E.; Woon, W.L.; Janajreh, I.; Madnick, S. A unified approach for taxonomy-based technology forecasting. In *Business Intelligence Applications and the Web: Models, Systems and Technologies*; IGI Global: Hershey, PA, USA, 2012; pp. 178–197.
53. Kelly, F.P. *Charging and Accounting for Bursty Connections*; Michigan Publishing, University of Michigan Library: Ann Arbor, MI, USA, 1996.
54. Manhaeve, R.; Dumancic, S.; Kimmig, A.; Demeester, T.; De Raedt, L. Deepproblog: Neural probabilistic logic programming. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2018; Volume 31.
55. Piringier, H.; Tominski, C.; Muigg, P.; Berger, W. A multi-threading architecture to support interactive visual exploration. *IEEE Trans. Vis. Comput. Graph.* **2009**, *15*, 1113–1120. [[CrossRef](#)]
56. Shahabuddin, J.; Chrungoo, A.; Gupta, V.; Juneja, S.; Kapoor, S.; Kumar, A. Stream-packing: Resource allocation in web server farms with a qos guarantee. In *Proceedings of the International Conference on High-Performance Computing, Hyderabad, India, 17–20 December 2001*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 182–191.
57. Walhovd, K.B.; Fjell, A.M.; Dale, A.M.; Fischl, B.; Quinn, B.T.; Makris, N.; Salat, D.; Reinvang, I. Regional cortical thickness matters in recall after months more than minutes. *Neuroimage* **2006**, *31*, 1343–1351. [[CrossRef](#)]
58. Van der Stoep, N.; Nijboer, T.C.; Van der Stigchel, S. Exogenous orienting of crossmodal attention in 3-D space: Support for a depth-aware crossmodal attentional system. *Psychon. Bull. Rev.* **2014**, *21*, 708–714. [[CrossRef](#)]
59. Atchley, P.; Kramer, A.F.; Andersen, G.J.; Theeuwes, J. Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychon. Bull. Rev.* **1997**, *4*, 524–529. [[CrossRef](#)]