Title Generation on COVID-19 Using Combinations of Extractive Summarization, the T5 Transformer, and BART

Daniel Gruhn, Long Truong, Deon Seyfi, Kalin Zaluzec

Abstract

For this project, we focused on different methods of title generation and summary generation. There were three main approaches we used for summarization. These approaches are extractive summarization, the T5 transformer, and the BART model. For title generation, we only used the latter two of these approaches. When training and testing our model, we tested different hyperparameters for each of these approaches. Our goal was to refine title generation by using different combinations of the above methods along with different sets of hyperparameters.

1 Introduction

Titles can be seen as the most important part of the document; it is the part seen by the most people.(Tullu, 2019) There have been many studies on how different characteristics in titles cause it to be viewed or cited more often. For example, short titles are more appealing to longer titles which could be seen as complex and boring.(Paiva et al., 2012) Unless the title can grab the reader's attention, it does not matter what the contents of it are. The most important things that titles need to do are convey the main topics, highlight the importance, be concise, and be attractive to readers.

Each of these four aspects contribute to the creation of successful titles in unique ways. Conveying the main topic is the most important aspect for catching the attention of the target audience.(Schmieder, 2018) Highlighting the importance lets the audience know the difference between one article and another on a similar topic. Being concise ensures that every word in the title is doing its job by drawing the reader in rather than pushing them away by being drawn-out and boring. Finally, being attractive to readers is the basic function of the title.

Summarization on its own can fulfill two of these aspects. Conveying main topics is guaranteed by the definition of a summary and highlighting the importance of a document is likely already accomplished by adequate summaries.(Sethi et al., 2016a) It also directly helps with the creation of a concise title by limiting the amount of text to work with to the most important parts. A standard summary, however, has too many words for a title. Since extractive summarization takes direct quotes from the text, it is very likely to have unnecessary text mixed in with the important text. Attracting readers can be seen as the greatest variable factor as it can be highly opinionated.

There are many types of titles that change based on the type of writing. Titles in literature have multiple different variants including clear titles, distinguished titles, descriptive titles, and symbolic titles. (Spero, 2018) Meanwhile, academic writing only has three types of titles: declarative, descriptive and interrogative.(Schmieder, 2018) Declarative titles are good because they get straight to the point and have a very clear final result presented. This makes it very easy to cite them by letting the audience know what kind of information it leads to without being too much of a time investment. Descriptive titles reference the methods used rather than the results directly. Interrogative titles ask the question that gets answered by the research. These questions can very easily draw attention, but since they have no details about how the research was conducted or the results, they is more difficult to deliberately search for using keywords.

This project explores different techniques for title generation. The specific technique that we will focus on is title generation based on summaries of a document rather than generating titles directly from a document. We plan to use varying approaches to summary generation and compare the resulting titles that we generate using these summaries as

input. We also plan to use more direct title generation methods that do not involve using summaries as a basis for comparison for our results.

2 Related Works

Title and summary generation are not a new topics in the field of NLP. There have been numerous different approaches to both of these subtopics. Two current widely used approaches to both of these NLP subtopics involve the use of the T5 transformer and the BART model. As these will be covered later in the paper, we will not talk about them in this section.

One example of title generation related to this project comes from a paper titled Automated Title Generation in English Language Using NLP. This paper outlines three different approaches to generating titles. (Sethi et al., 2016b) The first approach is the symbolic approach; this approach focuses solely on representing the knowledge from a given text in a title. The second approach is the statistical approach. This approach entails using mathematical techniques. It could also be described as a probabilistic implementation. Finally, there is the connectionist approach, which is a combination of the previous two approaches. In addition to these approaches, a few other techniques are outlined. One such technique is implementing a part-of-speech tagger. (Sethi et al., 2016b) This is a tool that allows for lexical analysis. Various part-of-speech tagging tools are readily available; some examples are the OpenNLP Tagger and the natural language toolkit (NLTK). One technique mentioned is called discourse analysis. (Sethi et al., 2016b) Discourse analysis is essentially taking all pronouns within a given text, and replacing them with their corresponding nouns. By doing so, it allows for our title generation models to better construct titles that are relevant to the input text. After this, we focused our attention on summarization. (Sethi et al., 2016b)

The core idea of this project is to use principles of extractive summarization to more methodically truncate our text in preparation for generating titles. To this end, we did a lot of research regarding extractive summarization. One topic in particular we read about was event-based extractive summarization. This method of generating summaries involves four main steps. (Vanderwende et al.) Firstly, the text is broken down into smaller units (sentences or paragraphs). Secondly, each of these smaller units will be tagged with a concept. Thirdly, units will be chosen based on their topics to become part of the final summary. Lastly, the third step will be repeated until the desired length is reached. The major challenge associated with this type of extractive summarization is feature recognition (determining the topic of each unit). (Vanderwende et al.) After this step is completed, generating an extractive summary can be done quite easily by choosing the units of text with features that show up most often. (Vanderwende et al.)

Another approach to title generation is called SummaRuNNer. SummaRuNNer is a recurrent neural network based sequence model. (Nallapati et al., 2016) The model is used to perform extractive summarization before generating summaries of input text. Rather than using the extractive summarization methods that we used for this project, SummaRuNNer uses sequence classification to determine which sentences from a given text are the most important. (Nallapati et al., 2016) Once the most important sentences are selected, SummaRuNNer will generate a summary using the selected sentencecs. (Nallapati et al., 2016) Despite this key difference, the design philosophy behind SummaRuNNer closely mirrors the approaches we used and was one of the inspirations for the direction we took this NLP project. (Nallapati et al., 2016)

3 Methods

When we approached the problem of summarization (and by extension, title-generation), we focused on two different methods. These methods are extractive summarization and abstractive summarization. Extractive summarization takes parts of the text that are deemed important word-for-word and uses them to construct a summary.(Sciforce, 2019) What changes in extractive summarization models is deciding which words are important enough to be kept and which are to be thrown out. Abstractive summarization attempts to find the general meaning of a given text and broadly explain the main idea while leaving out less crucial details. Due to the generation of new text as a summary, information may not necessarily be the same as the original document. For all abstractive summarization, we trained our models using the COVID-19 open research dataset challenge (CORD-19).

3.1 The CORD-19 Dataset

The CORD-19 dataset consists of nearly 360,000 files containing documenst of and relating to covid-19. The contents of these documents consisted of titles, abstracts, and main texts. Initially, we planned to experiment with using abstracts in our title generation, but due to time constraints, we felt that our efforts were better spent in other areas of the project. We chose this dataset because of the size as well as the documented academic titles and organized text to go along with the titles. This data set was not without its flaws, however. One potential issue with this dataset is the lack of diversity within it; all documents in this dataset were about covid-19 in some manner. Secondly, we had to filter some documents out of the dataset. Some documents were unusable for various reasons. One semi-regular issue was the occurrences of documents that were not in English. Another was formatting issues. The following figure is of a page of one of the problematic documents.

BMJ Open Page 14 of 27

DISCUSSION
In this study, we found that while SGMT has room to grow, the practice is leader in delivering culturally safe care in their community. We also found that Indigenous participants conceptualize cultural safety in distinct ways that reflect cross-cutting Indigenous social constructs and understandings of the interwoven nature of relationships, spaces, knowledge, and practice.[53-55] This provides further evidence that Indigenous values and approaches to health, wellbeing, and society persist in urban spaces, despite the colonial imposition of Western biomedicine.[3 7-9 11 12] In this discussion, we will explore Indigenous cultural safety and its connections to cultural expression, cultural continuity, and Indigenous resurgence. We will also explore the parallels between the Indigenous and racialized non-Indigenous participants' conceptualizations, and between all participants in terms of the anti-oppressive benefits of cultural safety.

This discussion draws on Indigenous perspectives that understand "culture" as integral to social structuring and relationships, representing "the interconnected social totality of the distinct mode of life encompassing the economic, political, spiritual, and social."[4 p65] Cultural expression and continuity, then, are as much about ceremonial knowledge and protocols as they are about values, ethics of relation, knowledge systems, and social and political structure. Additionally, while we use the collective terms "Indigenous" and "Indigenous-specific," the diversity and richness of First Nations, Inuit, and Métis cultural heritage and expression as Indigenous peoples from many nations and backgrounds coming together in a large urban centre should not be underestimated. Support for self-determined and relationship-based cultural expression, cultural respect, and sharing across these diversities is a core premise of SGMT's practice,[52] and cultural safety in Toronto more generally.

Indigenous conceptualizations of cultural safety
Although the Indigenous participants were from different communities, there were several elements of cultural safety that emerged as "Indigenous-specific," or nearly so. For example, expressing a shared desire for **two-way, kin-based relationships**, the Indigenous participants described their ideals and their experiences at SGMT as being "amongst sisters" who "get it," creating feelings of trust, connection, and confidence in the care they receive. This sense that family is the "foundational relationship for pursuing any economic, political, social, or cultural activities and alliances"[57 p433] resonates with Indigenous social constructs such as Cree and Métis conceptualizations of *wahkootowin*. Métis elder and scholar Maria Campbell explains wahkootowin, or the "kinship or the state of being related" (Ermine as cited in 58 p5], as follows:

> "Today it is translated to mean kinship, relationship, and family as in human family. But at one time, from our place it meant the whole of creation. And our teachings taught us that all of creation is related and inter-connected to all things within it. Wahkootowin meant honouring and respecting those relationships. [It was] our stories, songs, ceremonies, and dances that taught us from birth to death our responsibilities and reciprocal obligations to each other. Human to human, human to plants, human to animals, to the water and especially to the earth. And in turn all of creation had responsibilities and reciprocal obligations to us." (as cited in 58 p6)

The Indigenous participants' desire to connect with their midwives as family for the process of childbearing may be interpreted as an expression of wahkootowin, and/or the diverse yet resonant ways in which familial or kinship relationships remain foundational to social order across Indigenous societies.

For some Indigenous peoples, there is the notion that if we relocate ourselves within the networks of kinship such as wahkootowin, we can heal the ruptures of multi-generational family disruption, abuse,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml 13

BMJ Open: first published as 10.1136/bmjopen-2020-038168 on 29 September 2020. Downloaded from http://bmjopen.bmj.com/ on May 12, 2021 at Kaiser Consortia. Protected by copyright.

Figure 1: Example of an unusable document from the CORD-19 dataset (Churchill et al., 2020)

While the page in Figure 1 looks fine, the big issue with it are the line numbers and vertical/diagonal text. When read from our dataset, some parts of the document appeared as the following:

Despite the 24 fact that we did not find any association between comorbidities such as diabetes and cancer 25 and the mean death rates at country level, it is possible for an individual with any or all of 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 F o r p e e r r e v i e w o n l y 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59

Figure 2: Representation of data from Figure 1 as read from the JSON file

Because of the result in Figure 2 and documents that resulted in similar looking data, we had filter out the anything unusable before we did any training.

3.2 Weight-Based Extractive Summarization

The first thing we did when approaching this project was to code something to apply extractive summarization to text. We kept this to be sentence-based, so our extractive summarization methods were unable to create summaries shorter than a single sentence. With this in mind, we only used this procedure as a method of text truncation before applying other title generation methods rather than as a stand-alone method for generating titles.

Our extractive summarization method involves essentially counting words. For each document, we would count the number of occurrences for each word in the document. This is done to determine which words are the most important. In order to only have the meaningful words being marked as important, stop words (including words such as 'the', 'and', 'then', etc.) were omitted from this process. Then, for each sentence, we could calculate the weight of that sentence based on the total counts for each word in the sentence. The sentences were then sorted by weight, and the sentences with the highest weight were kept while the rest were discarded. This process (in theory) leads to only the most crucial sentences to the text remaining.

3.3 The T5 Transformer

The 'T5' in T5 transformer stands for 'Text-to-Text Transfer Transformer. While the focus of this report is document summarization, the T5 model

was trained to do a variety of other tasks including question answering, classification, and translation. The purpose of this is to allow the same model, loss function, and hyperparameters to be used across every task the model is designed for. (Raffel et al., 2020a) To differentiate different tasks from each other, a prefix is required on inputs. (Raffel et al., 2020b) For example, the prefix 'translate English to German: ' would be an appropriate prefix. For our purposes, we used the prefix 'summarize: ' to indicate that we desire the T5 transformer to generate a summary for us.

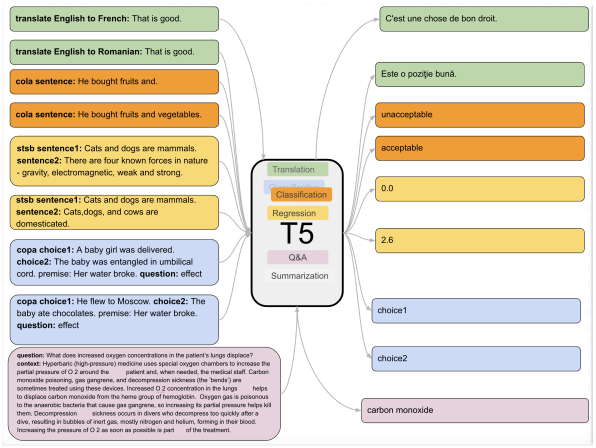

Figure 3: Example inputs and outputs of the T5 transformer (Rajasekharan, 2019)

The T5 model was trained and tested on the 'Colossal Clean Crawled Corpus' (C4). The cleaned (duplication removal, bad word removal, etc.) C4 dataset consists of millions of English text entries scrapped from around the internet. (Raffel et al., 2020a) One interesting fact to note is that the common crawl regularly produces about 20 terabytes of scraped text per month. The majority of this scraped text cannot be used for the dataset; it consists mostly of non-useable gibberish, text from menus, or error messages. (Raffel et al., 2020a) Other forms of text that are not useful are also found in this such as offensive text or source code text. The C4 dataset was cleaned according to the following criteria: (Raffel et al., 2020a)

1. Lines without a terminal punctuation mark were discarded

2. Pages with fewer than 5 sentences were discarded

3. Lines with fewer than 3 words were discarded

4. Any page that contains a word from the 'List of Dirty, Naughty, Obscene, or Otherwise Bad Words' was discarded (Emeric, 2020)

5. Any line that contains the word 'javascript' was discarded (to remove javascript errors)

6. Any page with the placeholder phrase 'lorem ipsum' was discarded

7. Any pages with a certain configuration of curly braces ('{' or '}') were discarded to remove code from the dataset

8. Any span of three sentences that occurred more than once were discarded

The T5 pretrained model comes with 5 different variants: t5-small, t5-base, t5-large, t5-3b, t5-11b. The t5-small model has 6 attention modules and the t5-base has 12 attention modules. All other models have 24 attention modules. Additionally, these models have 60 million, 220 million, 770 million, 3 billion, and 11 billion parameters respectively. (Raffel et al., 2020a) Each t5 model was evaluated using 34 different metrics (such as GLUE average). In multiple categories, the t5-11b model outperformed the previous best prior to the creation of the t5 model. In all cases, the t5-11b model outperformed the other variants of the t5 model.

3.4 The BART model

BART is a denoising autoencoder for training seq2seq models and has demonstrated that it can be effectively fine-tuned on tasks such as natural language generation, translation, comprehension, and especially summarization. (Lewis et al., 2019) BART model is a combination of bidirectional and auto-regressive transformers. This standard machine translation architecture of BART is a more general version of BERT model, where the encoder part corresponds to the structure of BERT (Devlin et al., 2018) and the decoder part is the auto-regressive left-to-right decoder following the settings of GPT. (Radford et al., 2018)

The model has some additional modifications from the original schematics of BERT and GPT. The model excludes BERT's feed-forward neural network before the word prediction portion. However, BART still contains approximately 10 percent more parameters compared to the equivalent-sized BERT model. Regarding the decoder stage, it replaces the non-linear activation function ReLU

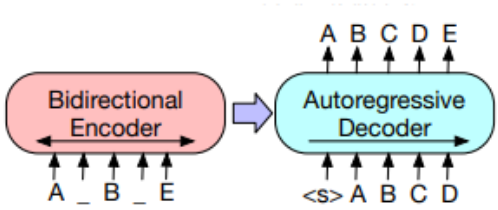

Figure 4: BART architecture (Devlin et al., 2018)

with GeLUs (Hendrycks and Gimpel, 2016) and additionally performs cross-attention over the final hidden layer resulting from the encoder. During the pre-training process, BART receives the corrupted document with an arbitrary noising function as input and performs the task of optimizing the cross-entropy to predict and reconstruct the original uncorrupted document. The noising transformations applied to the original sentences, which will be randomly shuffled during the pre-training process, are illustrated in Figure 5.

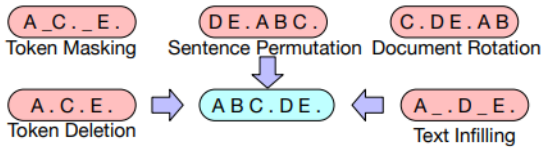

Figure 5: Variations of Noising Function of BART Applied to Input Sequence (Devlin et al., 2018)

Due to the autoregressive decoder implementation, BART can be directly fine-tuned for sequence generation tasks such as summarization. The information is copied from the input and manipulated slightly using the noising function. Then, the input sequence will be forwarded to the encoder and the decoder will be responsible for generating the output auto-regressively. This output is closely related to the denoising pre-training objective. As the result, the model can effectively learn the contextual representation of the input document. Specifically fine-tuning for the summarization task, the bidirectional encoder stage encodes the original document of body text, and the left-to-right decoder stage predicts the reference summary. In our experiment, we utilized the large model of BART (bart-large-cnn), which is composed of 12 layers where each encoder and decoder stage was initialized from the weights pre-trained on CNN/Daily Mail dataset.

3.5 Modeling

We implemented the T5 transformer and BART model using pre-trained models from natural language processing library Hugging Face. Specifically, we utilized T5ForConditionalGeneration with T5Tokenizer for the T5 model and BartForConditionalGeneration with BartTokenizer for the BART model. In order to avoid having problems with the encoder stage, we added additional arguments to tokenizer's encode function: max_length (512 for T5 and 1024 for BART) and truncation (True). The base hyperparameters utilized for the initial model architectures illustrated in the following table.

Variable	Value
data_size	10000
train_split	0.99
max_epochs	200
learning_rate	0.001
model_type	't5-small'
batch_size	1000
text_truncation	'no_truncation'

3.6 Training and Evaluation

When training our model, we conducted a grid search to find the best set of hyperparameters. Our grid search initially consisted of the following different hyperparameters:

Hyperparameter	Possible Values
Model Type	't5-small', 'BART'
Batch Size	100, 500, 1000
Truncation	100, 500, no truncation
Learning Rate	0.01, 0.001, 0.0001

This would have resulted in 54 different models being trained, which we felt would have been too time-consuming for us to do considering the impending deadline. With that in mind, we removed the learning rates from the set of hyperparameters. In the end, we trained 18 different models in our grid search.

Most of our hyperparameters are straightforward, but our truncation size hyperparameter is perhaps the most important one. The truncation size will dictate how we apply extractive text summarization to reduce the size of the text to either under 100 words, under 500 words, or not at all. In the case where we do not apply extractive summarization,

this does not mean that our text will not be truncated. It means that the tensors will be truncated according to the max input size (512 for T5, 1024 for BART) that can be accepted by the model.

In order to evaluate our models, we had no objective way to do this. Instead, we had all four of our team members evaluate titles. For each given document, we each chose the hyperparameter set that resulted in the title that we considered the best to give each hyperparameter a score. After doing so we normalized each set of hyperparameter scores to be between 0 and 100 to make the results easier to interpret.

4 Results

For our results, we decided to evaluate our results in two different sets. These two sets include all results produced by T5 and all results produced by BART.

4.1 T5 results

Using the T5 transformer, we had a degree of success in our title generation, but there is still room for improvement. The first thing we did with our results was determine which set of hyperparameters yielded the best titles. The following are the top three performing sets of hyperparameters pertaining to the t5-small model.

Hyperparameter Set	Normalized Score
Batch Size: 100 Truncation: 100	100
Batch Size: 1000 Truncation: 100	29.6
Batch Size: 1000 Truncation: 500	24.1

There were a few titles in particular that we thought were pretty good that came from our fine-tuned t5-small model. One such example is 'How the Field of Infectious Diseases Can Leverage Digital Strategy and Social Media Use During a Pandemic.' While we all agreed that this is a good title and that it matched the content of the paper, it is somewhat of an interesting result. The original title was 'Journal Pre-proof COVID-19 anti-vaccine movement and mental health: Challenges and A way forward Ramdas Ransing Title: COVID-19 anti-vaccine movement and mental health: Challenges and A way forward.' The original title doesn't talk about social media, yet our model shifted the focus from the anti-vaccine movement to social media.

We also had a set of bizarre results from the T5-transformer. Two titles generated from separate documents were 'Innovative liver research continues during the current pandemic' and 'Innovative liver research continues during the early COVID19 pandemic.' Given how similar these titles are, we might expect the input documents to be nearly the same. This was not the case, however. The original titles for these documents were 'Multi-Task Driven Explainable Diagnosis of COVID-19 using Chest X-ray Images' and 'Supporting Information Quantification of mRNA Expression Using Single-Molecule Nanopore Sensing' respectively. We weren't able to precisely determine why this happened, but we have had a few theories. The most likely things that we feel are to be the causes are using the t5-small model, having a data set that is too small, model overfitting, and using a data set that is not diverse enough.

4.2 BART results

Regarding the BART model, the model experienced unstable issues. This resulted in generally bad and repetitive titles. Variations of the BART model with different sets of hyperparameters were sometimes optimized down to 0.08 in negative log loss value. However, the generated results for different articles' titles were only the word "and." Retraining the same setup of models only converged to a loss value between 2 and 4. Yet, the generated titles were somewhat better and more relevant. The generated title for "Methods for Gene Delivery" is "rology States u United Cov perspectivesact of a to." This result title seemed to generate different tokens in the incorrect positions. We have tried to change Huggingface's default loss function of BART model negative log loss to binary cross-entropy loss function, yet the results did not improve. We were not able to figure out the portion in the code that might cause the issue.

5 Conclusion and Further Research

Our work on this project has produced mixed results. We are satisfied with the generated titles we were able to retrieve using the methods chosen for this project, but there is much more room for improvement. We are confident that with more time to work, the quality of the generated titles could be improved significantly.

Since the current sizes of the dataset trained on both T5 model and BART model were 1,000 (for basic tests) and 10,000 (for our final results), increasing the size of the data that we trained on would be one huge area for improvement. Training on 10,000 articles already took a large amount of time for our grid search, but if time and resources were to permit, these models would benefit from being trained on 100,000 documents or more. If given the time and resources, we would even consider using the entire dataset, which consists of over 300,000 documents. We believe that training on a large sample of the dataset would better represent the dataset as a whole, and would help us overcome the issue of our model overfitting the data. To further improve the performance of the models, we would like to include more hyperparameters aside from different batch sizes and text truncation lengths in our grid search process. Some of these hyperparameters would include learning rates, optimizers, and loss functions. In addition, different variants of the T5 transformer other than T5-small could be tested such as t5-11b and t5-large. Finally, we also want to train the models on datasets other than the current covid-19 dataset to determine if the (lack of) diversity within our data set affected the performance and results of our models.

In addition to these improvements in the methodology, we would like to improve our method of evaluating results. Currently, only our four members evaluated the titles. As much as we tried to be unbiased and objective, it would be dishonest to say that we were completely unbiased when it comes to evaluating our results. For a better and more unbiased evaluation of results, having more people unrelated to and unaware of the objective of this project to evaluate titles will be far more effective to our goal.

Overall, we feel that we produced satisfactory results for this project. If we were to continue working on the project and the above changes were made, we are confident that we could significantly improve the quality of the titles that we are able to generate.

References

Mackenzie Churchill, Janet St Michael's Hospital Centre for Urban Health Solutions, Well Living House Smylie, Centre for Research on Inner City Health Saint Michael's Hospital, Sara University of Toronto, Dalla Lana School of Public Health Wolfe, Cheryllee Seventh Generation Midwives Toronto, Bourgeois, Helle Seventh Generation Midwives Toronto Moeller, Michelle Lakehead University Firestone, and Centre for Research on Inner City Health St. Michaels Hospital. 2020. Conceptualizing cultural safety at an indigenous-focused midwifery practice in toronto, canada: Qualitative interviews with indigenous and non-indigenous clients.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jacob Emeric. 2020. List-of-dirty-naughty-obscene-and-otherwise-bad-words.

Dan Hendrycks and Kevin Gimpel. 2016. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *CoRR*, abs/1910.13461.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2016. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. *CoRR*, abs/1611.04230.

Carlos Eduardo Paiva, JoÃPaulo da Silveira Nogueira Lima, and Bianca Sakamoto Ribeiro Paiva. 2012. Articles with short titles describing the results are cited more often. *Clinics*, 67:509 – 513.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020a. Exploring the limits of transfer learning with a unified text-to-text transformer.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020b. T5.

Ajit Rajasekharan. 2019. T5 — a model that explores the limits of transfer learning.

Eric Schmieder. 2018. How to write an engaging title for your academic journal article.

Sciforce. 2019. Towards automatic text summarization: Extractive methods.

Nandini Sethi, Prateek Agrawal, Vishu Madaan, Sanjay Singh, and A. Kumar. 2016a. Automated title generation in english language using nlp. 9:5159–5168.

Nandini Sethi, Vishu Madaan, Prateek Agrawal, and Sanjay Kumar Singh. 2016b. Automated title generation in english language using nlp. *International Science Press*.

Joei Spero. 2018. The four types of titles: Which is best for your story?

Milind S. Tullu. 2019. Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. *Saudi journal of anaesthesia*.

Lucy Vanderwende, Michele Banko, and Arul Menezes. Event-centric summary generation.