Text summarization

Maya Angelova
mangelova@uni-potsdam.de

March 11, 2021

The topic of the current project is the Text Summarization Task. Different approaches were tested and implemented – two extractive methodologies – a naive word frequency based implementation and a TextRank based one. Further two abstractive methods were utilized – an implementation of an LSTM Encoder Decoder Neural Network with Attention and already pretrained Transformer models.
The best results according to automatic evaluation metrics are produced by the naive extractive approach with texts containing less than 300 words, the worst with the naive LSTM with Attention network.
The objective of the project was to get acquainted with different solution approaches for solving the task using different frameworks and libraries. I implemented the code and wrote the report. The main take-away messages from this project are that in order to get good results implementing a simple neural network from scratch is not always the best approach if there are not enough data and resources. Powerful models such as BART or T5 can be implemented and trained or used out of the box when pretrained on the specific task. Simple extractive approaches produce good enough results for a comprehensible summary.

1 Overview

The Text Summarization Task is generally described as converting long texts into short ones or in the words of Jacob Eisenstein:

"Text summarization is the problem of converting a longer text into a shorter one, while still conveying the key facts, events, ideas, and sentiments from the original." [6]

There are several ways to perform text summarization conditional on the number of input documents (single or multiple), the purpose (domain specific, general or query-based) and/or thee output generation (extractive or abstractive). [1]
The two main approaches towards this task are the **extractive** where "the summary is a subset of the original text [..] choosing a subset of the document that best covers the

concepts mentioned in the document as a whole“ [6] and the **abstractive** one where the summary is “produced de novo, by paraphrasing the original“ [6]).
In this project the summarization task is performed each time on a single document, an article from the WikiHow dataset [7], using both extractive and abstractive approach implementations. The following is a report of the methodology and the results.

2 Data

I chose the WikiHow dataset: *WikiHow: A Large Scale Text Summarization Dataset* [9], obtained from [7]. This dataset contains around 230,000 article and summary pairs, taken from the online WikiHow[1] knowledge base. The articles are written by different human authors having diverse topics. Each article comprises multiple paragraphs which in turn start with a summary. Paragraphs and summary sentences form pairs which are used to train and/or evaluate summarization methods.

Preparation First, the title, text and headline fields of each article are pre-processed.
The implemented data preprocessing functionality are slighly different for each of the two approaches – extractive and abstractive. The shared preparation steps were amongst others: remove all the duplicate rows and those containing NULL values for the text and headline columns; convert to lowercase and to plain ASCII, apply contraction conversion and further remove digits at the end of the line, new line characters, url markers, etc. For the abstractive approach the punctuation is further removed from the text as well as other minor cleansing steps.
The data is divided into short and shorter articles, where the former contain all the text with lesser than 800 words and the latter with less than 300.

3 Extractive Summarization

Extractive summarization generates the text as “a subset of the original text [..] choosing a subset of the document that best covers the concepts mentioned in the document as a whole“ [6], meaning that there is no novel text generation. One of the advantages of such methods is the correct usage of grammar and syntax, preconditioned on the input document and thus do not require extensive linguistic analysis, as well as no additional training time and resources as with neural networks.
The main question of how to pick the most important and representative sentences of the text can have different solutions. Two of them are presented further on.

[1] http://www.wikihow.com/

3.1 Word Frequency Naive Approach

One such approach would be using features as word and phrase frequency to detect and extract key sections from the text. This dates back to 1950 when Luhn [2] suggests ranking each sentences according to high frequency words while ignoring the high frequent and very common words and is the inspiration of the current implementation.
The implemented naive methodology includes tokenizing each document, building a vocabulary and counting the occurence of each word in it, while removing stop words (high frequency common words) from the document, so they do not introduce noise into the applied summarizing function. After computing the relevant frequencies for each word in the document, each sentence is rated according to significant terms it contains and assigned an importance score. Sentence extraction is performed based on the score by selecting the top k (here the top 3) most significant sentences as the summary.

3.2 TextRank Extractive Approach

TextRank is based on Google's PageRank algorithm, which in turn computes the probability of a user visiting a page. TextRank [13] is graph-based and is used to find the most relevant sentences and keywords, sort them and generate an automatic summary from the k-best ranked ones. A brief description of the algorithm is as follows: every parsed sentence is a node in a graph, where the edges contain the similarity score (in this implementation a cosine similarity) for two sentences (nodes). After the construction of the graph, the PageRank algorithm is applied to it. The main assumption is that the most important sentence will be the one that is the most similar to the rest of the document.

4 Abstractive Summarization

Abstractive summarization generates the text "de novo, by paraphrasing the original" [6], where new words can also be generated, paraphrasing and thus mimicing human cognition.

4.1 Naive Long Short-Term Memory Neural Network with Attention

The abstractive summarization approach according to [6] can be viewed as a translation problem. For such class of tasks an Encoder Decoder neural network (or sequence to sequence [15])) can be applied. The Encoder Decoder architecture with recurrent neural networks reads the whole input (sentence, text, word), condenses its representation to an internal fixed-length hidden state, which is then fed into the decoder to produce words/characters till an end indicating token is reached. To mitigate the problem stemming from long input sequences loosing information from their first tokens, an attention mechanism is applied, allowing the decoder to look at any particular state of the encoder and thus learn and focus on the relevant parts in order to produce the current output.

I implemented such a model as described by [6], where each word in the text is represented by an one-hot vector and the enoded text acts as the input to the encoder. The decoder takes the hidden state of the encoder, applies an additive attention mechanism [14], focusing on the different parts of the input data and produces a new sequence – the summary.

4.2 Pretrained Neural Network Models

As further experiments, two pretrained neural network models from the summarization pipeline provided by huggingface[2] were utilized and tested: BART (*Bidirectional and Auto-Regressive Transformers* [16]) and T5 (*Text-to-Text Transfer Transformer* as presented in *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer* [11]). Both models are based on the Transformer Encoder Decoder architecture, allowing for parallelization of processing the input sequence and solving the long-range dependency more efficiently. The main objective of T5[3] is handling the input and the output as text. In comparison to the BERT architecture T5 is using a casual decoder and additional alternative pre-training tasks instead of the original cloze task. The BART architecture combines both bidirectional and auto-regressive Transformers – a BERT-like encoder with a GPT-like decoder[4].

5 Evaluation

The applied automatic evaluation metrics used are ROUGE_L [24], BLUE (1-4) [10], METEOR [25]. These were computed using the nlg-eval[5] python module. Tables with listed evaluation results are to be found in the Appendix.
Bilingual Evaluation Understudy Score (BLEU) computes a score for the similarity of a generated sentences to a reference sentence with a score point between 0 and 1, with 1 indicating a perfect match. The calculation is based on counting n-grams overlaps and measures precision. The METEOR metric is a weighted F-measure, a combination of recall and precision based on mapping unigrams and penalyzing incorrect word order [6]. Recall Oriented Understudy for Gisting Evaluation (ROUGE) is a family of metrics based on recall calculation. In this project the ROUGE-L one is being reported, which is computed based on the longest common subsequence.
TextRank summarized 77436 texts containing less than 800 words, and 32725 texts containing less than 300 words. Naive Frequency approach summarized 178953 texts containing less than 800 words, and 103939 texts containing less than 300 words.
The LSTM model with Attention is trained on 75115 using teacher forcing and evaluated on 25039 texts. BART model summarized 5647 texts containing less than 800 words amd

[2]https://huggingface.co/transformers/main_classes/pipelines.html#transformers.SummarizationPipeline
[3]https://huggingface.co/transformers/model_doc/t5.html
[4]https://huggingface.co/transformers/model_doc/bart.html
[5]https://github.com/Maluuba/nlg-eval.git

T5 summarized 2527 texts containing less than 300 words.
As per Table 1 depicting the extractive summarization results, the shorter text summaries yield a slightly better performance in general, on the BLEU and ROUGE metrics (but not on METEOR). From the almost identical evaluation outcomes one cannot conclude which extractive methodology is the better one.
As of the abstractive approach the best performance is given by the T5 model summarizing 300 word texts, as can be seen in Table 3. Those powerfull pretrained models though summarizating unknown data perform better due to the volume of their training data, model parameters and architecture and resources used for training.
The naive implementation of LSTM network generates gibberish, due to the shallow network, the unsifficient time for training as well as unreliable resources to train for longer time (Google Colab with a time and usage limit). This renders this implementation attempt obsolete for purposes other than understanding and prorgamming practice.
Examples of original texts, original summaries and generated summaries for the different methodologies are given in the Appendix.

6 Outlook

There are many possible imporvements for the presented methodologies in this project.
Regarding the extractive method naive word frequency indicator representation approach introduced earlier such may include additional features such as sentence length, location in the text, cue words such as "important" etc. or topic related phrases. Different similarity computation functions with the TextRank algorithm should produce different results and is of interest to see different outcomes.
The drawbacks of the naive LSTM implementaion with Attention and its poor results are evident: not enough data, time and resources for training, simple one-hot vector embeddings. Possible improvements might include experimenting with different cells such as GRU, number hidden of layers and units, different regularization parameters, pretrained vector embeddings such as GloVe[6] or from projects such as GPT-2[7]. Transfer learning is also very interesting road to explore as well.
Evaluation is done using automatically computed metrics. These are problematic because they do not directly consider sentence structure or handle morphologically rich languages well, it is easy for an summarization algorithm to use synonyms and retain meaning but yield a low automatic evaluation score. Some mitigation strategies include human evaluation, though time consuming and expensive or a learned evaluation metric based on BERT and trained on a public collection of ratings based on the WMT Metrics Task dataset BLEURT [23] which can also be utilized.

[6] https://nlp.stanford.edu/projects/glove/
[7] https://openai.com/blog/better-language-models/

7 Acknowledgement

The following blogs and tutorials were inspirational for the implemenation of this project: *Automatic Text Summarization Made Simple with Python* [3], *Text summarization using TextRank in NLP* [4], *Practical NLP: Summarising Short and Long Speeches With Hugging Face's Pipeline* [5], *NLP From Scratch: Translation with a Sequence to Sequence Network and Attention* [17].
The frameworks and libraries used for the implementation include PyTorch [17], NumPy [18], NLTK [19], Scikit-learn [20], Spacy [21] and pandas [22]. The experiments were provided as Jupyter Notebooks in a GitHub repository[8].

8 Disclaimer

Errata include the following commits:

- The LSTM Notebook was not commited properly initially (ec68af2a12c51b002f3cc99d7543c979e55f8b5a)
- In the extractive summarization naive frequency implementation example, not the expected english stop words where applied but an non existent set which defaulted to the portuguese ones (cb7a0183d8e9f9f7837d4a03a49a5fc94485298a)
- In the extractive summarization TextRank implementation example, there was a more appropriate calling of the PageRank graph algorithm method (554e819e4d9e64d5d26300fbb1009a6c7

Maya Angelova

[8] https://github.com/mayaang/anlp

A Results

A.1 Extractive Approach Results

Metric	Naive < 800 words	Naive approach < 300 words	Text Rank < 800 words	Text Rank < 300 words
BLEU (1)	0.185197	0.206699	0.185741	0.192347
BLEU (2)	0.087741	0.098721	0.081665	0.086610
BLEU (3)	0.043440	0.049261	0.038681	0.041719
BLEU (4)	0.023255	0.026557	0.020044	0.021832
METEOR	0.133666	0.122656	0.114595	0.114719
ROUGE_L	0.144828	0.146027	0.133943	0.134614

Table 1: Results of text summarization with extractive methods: TextRank and Naive Frequency Approach (using spanish stop words for filtering)

Metric	Naive < 800 words	Naive approach < 300 words
BLEU (1)	0.185575	0.205678
BLEU (2)	0.087310	0.097544
BLEU (3)	0.043255	0.048618
BLEU (4)	0.023112	0.026142
METEOR	0.134472	0.123495
ROUGE_L	0.143539	0.144765

Table 2: Results of text summarization with Naive Frequency Approach using english stop words for filtering.

A.2 Abstractive Approach

Metric	LSTM with Attention (300 words)	BART (800 words)	T5 (300 words)
BLEU (1)	0.009899	0.187090	0.196875
BLEU (2)	0.000398	0.088135	0.096427
BLEU (3)	0.000000	0.044548	0.049443
BLEU (4)	0.000000	0.023773	0.026720
METEOR	0.009912	0.121261	0.117138
ROUGE_L	0.011821	0.146736	0.152443

Table 3: Results of text summarization with Neural Network models: LSTM Encoder-Decoder with Attention, BART, T5

B Summary: Naive Extractive Approach Short Example with up to 800 words

Remark: Correcting for the stop words produced total different results, as may be seen from this example:

B.1 Original Summary

keep related supplies in the same area., make an effort to clean a dedicated workspace after every session., place loose supplies in large, clearly visible containers., use clotheslines and clips to hang sketches, photos, and reference material., use every inch of the room for storage, especially vertical space., use chalkboard paint to make space for drafting ideas right on the walls., purchase a label maker to make your organization strategy semi-permanent., make a habit of throwing out old, excess, or useless stuff each month.

B.2 Original Text

if you are a photographer, keep all the necessary lens, cords, and batteries in the same quadrant of your home or studio. paints should be kept with brushes, cleaner, and canvas, print supplies should be by the ink, etc. make broader groups and areas for your supplies to make finding them easier, limiting your search to a much smaller area. some ideas include: essential supplies area – the things you use every day. inspiration and reference area. dedicated work area . infrequent or secondary supplies area, tucked out of the way.; , this does not mean cleaning the entire studio, it just means keeping the area immediately around the desk, easel, pottery wheel, etc. clean each night. discard trash or unnecessary materials and wipe down dirty surfaces. endeavor to leave the workspace in a way that you can sit down the next day and start working immediately, without having to do any work or tidying. even if the rest of your studio is a bit disorganized, an organized workspace will help you get down to business every time you want to make art. , as visual people, a lot of artist clutter comes from a desire to keep track of supplies visually instead of tucked out of sight. by using jars, old glasses, vases, and cheap, clear plastic drawers, you can keep things in sight without leaving it strewn about haphazardly. some ideas, beyond those just mentioned, include: canvas shoe racks on the back of the door wine racks with cups in each slot to hold pens/pencils. plastic restaurant squirt bottles for paint, pigment, etc., simply string up the wires across a wall or along the ceiling and use them to hold essential papers that you do not want to cut or ruin with tacks or tape. cheap and easy, this is also a good way to handle papers and ideas you touch regularly or need to pin up and down for inspiration., shelving is an artist best friend and is a cheap and easy way to get more room in your studio or art space. do not be afraid to get up high either, especially for infrequently used supplies. the upper reaches of the room are often the most under-utilized, but provide vital space for all your tools and materials., turning one wall into a chalkboard gives you a

perfect space for ideas, sketches, and planning without requiring extra equipment or space. you can even use it for smaller areas. paint over jars or storage equipment, allowing you to relabel them with chalk as your needs change. , a lot of disorganization comes when you keep moving the location of things, trying to optimize your space by reorganizing frequently. this usually has the opposite effect, leading to lost items and uncertainty when cleaning, but an afternoon with a label maker can solve everything. instead of spending all of your mental energy looking for or storing things, you can just follow the labels, freeing your mind to think about art., once a month, do a purge of your studio. if it is not essential or part of a project, either throw it out or file it away for later. artists are constantly making new things, experimenting, and making a mess. this is a good thing, but only if you set aside time to declutter. it may not be fun at the moment, but it is a lot more fun than spending minutes digging through junk to find the right paint or an old sketch. don't be sentimental here. if you have not used it in the last six months there is little chance you will use it in the next six months. toss it.

B.3 Generated Summary Corrected

some ideas include: essential supplies area – the things you use every day.by using jars, old glasses, vases, and cheap, clear plastic drawers, you can keep things in sight without leaving it strewn about haphazardly., shelving is an artist best friend and is a cheap and easy way to get more room in your studio or art space.

C Summary: Naive Extractive Approach Shorter Example with up to 300 words

C.1 Original Text

; , professional quality recordings of your songs are always preferred. , record a music demo. the word demo is shortened from the word demonstration. a demo is used to demonstrate your artistic abilities to the a and r representatives of record labels. , a demo submission is the process where an artist will find a record label company and send out their best three to five songs (demo) in the hopes of getting signed (receiving a record deal after signing a recording contract). , talk intelligently about your music, goals and where you fit with the label. ,,,

C.2 Original Summary

practice your material until you can perform each song perfectly., choose whether you are going to record your music in your own home studio or a professional recording studio., prove your talent., pick your best three to five songs for a demo submission., write a brief bio., visit for up-to-date information on how to properly submit your music demo. , complete the demo submission process. , wait a for a response.

C.3 Generated Summary Corrected

, record a music demo.the word demo is shortened from the word demonstration.a demo is used to demonstrate your artistic abilities to the a and r representatives of record labels., a demo submission is the process where an artist will find a record label company and send out their best three to five songs (demo) in the hopes of getting signed (receiving a record deal after signing a recording contract).

C.4 Generated Summary Erroneous

the word demo is shortened from the word demonstration.a demo is used to demonstrate your artistic abilities to the a and r representatives of record labels., a demo submission is the process where an artist will find a record label company and send out their best three to five songs (demo) in the hopes of getting signed (receiving a record deal after signing a recording contract).

D Summary: TextRank Extractive Approach Short Example with up to 800 words

D.1 Original Text

if you are a photographer, keep all the necessary lens, cords, and batteries in the same quadrant of your home or studio. paints should be kept with brushes, cleaner, and canvas, print supplies should be by the ink, etc. make broader groups and areas for your supplies to make finding them easier, limiting your search to a much smaller area. some ideas include: essential supplies area – the things you use every day. inspiration and reference area. dedicated work area . infrequent or secondary supplies area, tucked out of the way.; , this does not mean cleaning the entire studio, it just means keeping the area immediately around the desk, easel, pottery wheel, etc. clean each night. discard trash or unnecessary materials and wipe down dirty surfaces. endeavor to leave the workspace in a way that you can sit down the next day and start working immediately, without having to do any work or tidying. even if the rest of your studio is a bit disorganized, an organized workspace will help you get down to business every time you want to make art. , as visual people, a lot of artist clutter comes from a desire to keep track of supplies visually instead of tucked out of sight. by using jars, old glasses, vases, and cheap, clear plastic drawers, you can keep things in sight without leaving it strewn about haphazardly. some ideas, beyond those just mentioned, include: canvas shoe racks on the back of the door wine racks with cups in each slot to hold pens/pencils. plastic restaurant squirt bottles for paint, pigment, etc., simply string up the wires across a wall or along the ceiling and use them to hold essential papers that you do not want to cut or ruin with tacks or tape. cheap and easy, this is also a good way to handle papers and ideas you touch regularly or need to pin up and down for inspiration., shelving is an artist best friend and is a cheap and easy way to get more room in your studio or art space. do not be afraid to get up high either, especially for infrequently used supplies. the upper reaches of the room are often the most under-utilized, but provide vital space for all your tools and materials., turning one wall into a chalkboard gives you a perfect space for ideas, sketches, and planning without requiring extra equipment or space. you can even use it for smaller areas. paint over jars or storage equipment, allowing you to relabel them with chalk as your needs change. , a lot of disorganization comes when you keep moving the location of things, trying to optimize your space by reorganizing frequently. this usually has the opposite effect, leading to lost items and uncertainty when cleaning, but an afternoon with a label maker can solve everything. instead of spending all of your mental energy looking for or storing things, you can just follow the labels, freeing your mind to think about art., once a month, do a purge of your studio. if it is not essential or part of a project, either throw it out or file it away for later. artists are constantly making new things, experimenting, and making a mess. this is a good thing, but only if you set aside time to declutter. it may not be fun at the moment, but it is a lot more fun than spending minutes digging through junk to find the right paint or an old sketch. don't be sentimental

here. if you have not used it in the last six months there is little chance you will use it in the next six months. toss it.

D.2 Original Summary

keep related supplies in the same area., make an effort to clean a dedicated workspace after every session., place loose supplies in large, clearly visible containers., use clotheslines and clips to hang sketches, photos, and reference material., use every inch of the room for storage, especially vertical space., use chalkboard paint to make space for drafting ideas right on the walls., purchase a label maker to make your organization strategy semi-permanent., make a habit of throwing out old, excess, or useless stuff each month.

D.3 Generated Summary

paints should be kept with brushes, cleaner, and canvas, print supplies should be by the ink, etc.some ideas include: essential supplies area – the things you use every day.this usually has the opposite effect, leading to lost items and uncertainty when cleaning, but an afternoon with a label maker can solve everything.if you have not used it in the last six months there is little chance you will use it in the next six months.

E TextRank Extractive Approach Short Example with up to 300 words

E.1 Original Text

; , professional quality recordings of your songs are always preferred. , record a music demo. the word demo is shortened from the word demonstration. a demo is used to demonstrate your artistic abilities to the a and r representatives of record labels. , a demo submission is the process where an artist will find a record label company and send out their best three to five songs (demo) in the hopes of getting signed (receiving a record deal after signing a recording contract). , talk intelligently about your music, goals and where you fit with the label. ,,,

E.2 Original Summary

practice your material until you can perform each song perfectly., choose whether you are going to record your music in your own home studio or a professional recording studio., prove your talent., pick your best three to five songs for a demo submission., write a brief bio., visit for up-to-date information on how to properly submit your music demo. , complete the demo submission process. , wait a for a response.

E.3 Generated Summary

practice your material until you can perform each song perfectly., choose whether you are going to record your music in your own home studio or a professional recording studio., prove your talent., pick your best three to five songs for a demo submission., write a brief bio., visit for up-to-date information on how to properly submit your music demo. , complete the demo submission process. , wait a for a response.

F Summary: Neural Network LSTM Example

F.1 Original Text

fit a water balloon over a bottle of washable paint turn the bottle over and squirt some paint into the balloon do not fill the entire balloon with paint however you need room for the water liquid watercolor acrylic tempera and poster paint will all work great this method will give you a watered down tie dye effect tip the bottle right side up and pull the balloon off stretch the end of the balloon over a faucet or hose hold the balloon by the bottom then turn the water on let the balloon fill slowly then turn the faucet off pinch the end of the balloon and pull it off of the faucet place the balloon down into the sink then carefully knot the end fill each balloon with a different color of paint experiment with using different amounts of paint and water you can also shake the balloons to dilute the paint further keep the balloons in a bucket or plastic bin be careful not to pop them put on a set of clothes you do not mind getting dirty head outside where the paint will not ruin anything tape a sheet of poster paper against a wall or fence you can also prop a foam cord board up against a wall fence instead if you are not allowed to use the wall or fence spread the paper on the ground and stand up on a bench or chair the balloons will burst open when they hit the paper and splatter it with paint the water will mix with the paint and create a cool tie dye or watercolor effect if you are standing on a bench or chair simply drop the balloon onto the paper

F.2 Original Summary

squirt some washable paint into a water balloon fill the balloon the rest of the way with water knot the end of the balloon make more paint filled water balloons get ready to get dirty throw the balloons at the paper

F.3 Generated Summary

ribbon picked web interruptions closing player tails that reason staple that com gloss web account left closing player settings recent source closing naturally that detailed left com website that reason offers ¡EOS¿

G Summary: Neural Network BART Example

G.1 Original Summary

Keep related supplies in the same area., Make an effort to clean a dedicated workspace after every session., Place loose supplies in large, clearly visible containers., Use clotheslines and clips to hang sketches, photos, and reference material., Use every inch of the room for storage, especially vertical space., Use chalkboard paint to make space for drafting ideas right on the walls., Purchase a label maker to make your organization strategy semi-permanent., Make a habit of throwing out old, excess, or useless stuff each month.

G.2 Original Raw Text

If you're a photographer, keep all the necessary lens, cords, and batteries in the same quadrant of your home or studio. Paints should be kept with brushes, cleaner, and canvas, print supplies should be by the ink, etc. Make broader groups and areas for your supplies to make finding them easier, limiting your search to a much smaller area. Some ideas include: Essential supplies area – the things you use every day. Inspiration and reference area. Dedicated work area . Infrequent or secondary supplies area, tucked out of the way.; , This doesn't mean cleaning the entire studio, it just means keeping the area immediately around the desk, easel, pottery wheel, etc. clean each night. Discard trash or unnecessary materials and wipe down dirty surfaces. Endeavor to leave the workspace in a way that you can sit down the next day and start working immediately, without having to do any work or tidying. Even if the rest of your studio is a bit disorganized, an organized workspace will help you get down to business every time you want to make art. , As visual people, a lot of artist clutter comes from a desire to keep track of supplies visually instead of tucked out of sight. By using jars, old glasses, vases, and cheap, clear plastic drawers, you can keep things in sight without leaving it strewn about haphazardly. Some ideas, beyond those just mentioned, include: Canvas shoe racks on the back of the door Wine racks with cups in each slot to hold pens/pencils. Plastic restaurant squirt bottles for paint, pigment, etc., Simply string up the wires across a wall or along the ceiling and use them to hold essential papers that you don't want to cut or ruin with tacks or tape. Cheap and easy, this is also a good way to handle papers and ideas you touch regularly or need to pin up and down for inspiration., Shelving is an artist's best friend and is a cheap and easy way to get more room in your studio or art space. Don't be afraid to get up high either, especially for infrequently used supplies. The upper reaches of the room are often the most under-utilized, but provide vital space for all your tools and materials., Turning one wall into a chalkboard gives you a perfect space for ideas, sketches, and planning without requiring extra equipment or space. You can even use it for smaller areas. Paint over jars or storage equipment, allowing you to relabel them with chalk as your needs change. , A lot of disorganization comes when you keep moving the location of things, trying to optimize your space by reorganizing frequently. This usually has the opposite effect, leading to lost items and uncertainty when cleaning,

but an afternoon with a label maker can solve everything. Instead of spending all of your mental energy looking for or storing things, you can just follow the labels, freeing your mind to think about art., Once a month, do a purge of your studio. If it isn't essential or part of a project, either throw it out or file it away for later. Artists are constantly making new things, experimenting, and making a mess. This is a good thing, but only if you set aside time to declutter. It may not be fun at the moment, but it is a lot more fun than spending 30 minutes digging through junk to find the right paint or an old sketch. Don't be sentimental here. If you haven't used it in the last six months there is little chance you'll use it in the next six months. Toss it.

G.3 Generated Summary

An organized workspace will help you get down to business every time you want to make art as visual people . A lot of artist clutter comes from a desire to keep track of supplies visually instead of tucked out of sight by using jars old glasses vases and cheap clear plastic drawers you can keep things in sight without leaving it strewn about haphazardly .

H Summary: Neural Network T5 Example

H.1 Original Raw Text

practice your material until you can perform each song perfectly., choose whether you are going to record your music in your own home studio or a professional recording studio., prove your talent., pick your best three to five songs for a demo submission., write a brief bio., visit for up-to-date information on how to properly submit your music demo. , complete the demo submission process. , wait a for a response.

H.2 Original Raw Summary

Practice your material until you can perform each song perfectly., Choose whether you're going to record your music in your own home studio or a professional recording studio., Prove your talent., Pick your best three to five songs for a demo submission., Write a brief bio., Visit http://www.PhantomCityStudio.com/Demos for up-to-date information on how to properly submit your music demo. Complete the demo submission process. Wait a for a response.

H.3 Generated Summary

the word demo is shortened from the word demonstration a music demo is the process where an artist will find a record label company and send out their best three to five songs demo in the hopes of getting signed receiving a recording contract . a demo is used to demonstrate your artistic abilities to the a and r representatives of record labels .

References

[1] https://medium.com/sciforce/towards-automatic-text-summarization-extractive-methods-e8439cd54715

[2] Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of research and development, 2(2), 159-165.

[3] Automatic Text Summarization Made Simple with Python, https://medium.com/luisfredgs/automatic-text-summarization-made-simple-with-python-f9c3c645e34a

[4] Text summarization using TextRank in NLP, https://medium.com/data-science-in-your-pocket/text-summarization-using-textrank-in-nlp-4bce52c5b390

[5] https://towardsdatascience.com/practical-nlp-summarising-short-and-long-speeches-with-hugging-faces-pipeline-bc7df76bd366

[6] Jacob Eisenstein, Natural Language Processing

[7] Wikihow Dataset Download, https://github.com/mahnazkoupaee/WikiHow-Dataset

[8] NLP From Scratch: Translation with a Sequence to Sequence Network and Attention, https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html

[9] M. Koupaee and William Yang Wang, *WikiHow: A Large Scale Text Summarization Dataset*, https://arxiv.org/abs/1810.09305

[10] Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu (2002). Bleu: a method for automatic evaluation of machine translation. See acl (2002), pp. 311–318.

[11] Colin Raffel, Noam Shazeer, Adam Roberts, Kather-ine Lee, Sharan Narang, Michael Matena, YanqiZhou, Wei Li, and Peter J Liu. 2019. Explor-ing the limits of transfer learning with a uni-fied text-to-text transformer.arXiv preprintarXiv:1910.10683.

[12] NLP From Scratch: Translation with a Sequence to Sequence Network and Attention, https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html

[13] Mihalcea, Rada and Tarau, Paul, TextRank: Bringing Order into Text, https://web.eecs.umich.edu/ mihalcea/papers/mihalcea.emnlp04.pdf

[14] Dzmitry Bahdanau and Kyunghyun Cho and Yoshua Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, https://arxiv.org/abs/1409.0473

[15] Ilya Sutskever and Oriol Vinyals and Quoc V. Le, Sequence to Sequence Learning with Neural Networks, https://arxiv.org/abs/1409.3215

[16] M. Lewis and Yinhan Liu and Naman Goyal and Marjan Ghazvininejad and A. Mohamed and Omer Levy and Ves Stoyanov and Luke Zettlemoyer, BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, https://arxiv.org/abs/1910.13461

[17] PyTorch, pytorch.org/

[18] NumPy, numpy.org/

[19] NLTK, nltk.org

[20] Scikit-learn, https://scikit-learn.org/

[21] Spacy, https://spacy.io

[22] Pandas, https://pandas.pydata.org/

[23] Thibault Sellam and Dipanjan Das and Ankur P. Parikh, BLEURT: Learning Robust Metrics for Text Generation, https://arxiv.org/abs/2004.04696

[24] Lin, Chin-Yew, ROUGE: A Package for Automatic Evaluation of Summaries, https://www.aclweb.org/anthology/W04-1013

[25] Michael Denkowski and Alon Lavie, Meteor Universal: Language Specific Translation Evaluation for Any Target Language, https://www.cs.cmu.edu/ alavie/METEOR/pdf/meteor-1.5.pdf