Luka Žontar

Prilagoditev besedil kontekstu objave

DIPLOMSKO DELO

UNIVERZITETNI ŠTIŠIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: prof. dr. Zoran Bosnić

Ljubljana, 2020
This work is offered under the license Creative Commons Attribution-Sharing under the same conditions 2.5 Slovenia (or newer version). This means that the texts, pictures, graphs and other components of the work as well as the results of the diploma thesis can be freely distributed, reproduced, used, communicated to the public and processed, if the author and the title of this work are clearly and visibly indicated and in the case of modification, transformation or use by another work can be distributed only under the identical license. License details are available on the website creativewcommons.si or at the Institute for Intellectual Property, Streliška 1, 1000 Ljubljana.

The source code, its results and the developed software is offered under the licence GNU General Public Licence, version 3 (or later). This means that it can be freely distributed and/or processed under its terms. License details are available on the website http://www.gnu.org/licenses/.

This text is formatted with the editor $\LaTeX$. 
Faculty of Computer and Information science issues the following thesis:

When publishing a text, the writer has to consider who the target audience is. This is why it is important to adapt texts to different publication types (e.g. online newspapers, social media posts, official statements, research articles). The goal of this thesis will be to develop a methodology that can adapt texts (in English or Slovene language) to different profiles. The method will use summarization, text generation and paraphrasing to adjust the main text characteristics to appropriate target audience.
Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

My sincere thanks go to all the professors, who are continuously trying to improve our educational system, including my mentor, prof.dr. Zoran Bosnić, to whom I also owe gratitude for his patience and careful reviews that vastly contributed to the success of this thesis.

I would also like to thank Sarah, her mother and my family for standing by my side whenever I needed them.
Dedicated to Sarah, whose guidance helped me find myself.
Contents

Abstract

Izvleček

Razširjeni povzetek

1 Introduction
   1.1 Motivation .............................................. 1
   1.2 Goals and contributions ............................... 2
   1.3 Structure ................................................ 4

2 Related work
   2.1 Text evaluation metrics ................................. 5
   2.2 Length manipulation .................................... 9
   2.3 Paraphrase generation ................................. 11

3 Text evaluation metrics
   3.1 Structure ................................................ 13
   3.2 Publication types ....................................... 14
   3.3 Used metrics ........................................... 15
   3.4 Findings and interpretation .......................... 17

4 Length manipulation
   4.1 Structure ................................................ 22
   4.2 Extractive summarization .............................. 22
## List of acronyms used

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WNLL</td>
<td>word negative log-likelihood</td>
</tr>
<tr>
<td>RSRS</td>
<td>ranked sentence readability score</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>bidirectional long short-term memory</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>HAN</td>
<td>hierarchical attention networks</td>
</tr>
<tr>
<td>BERT</td>
<td>bidirectional encoder representations from transformers</td>
</tr>
<tr>
<td>WHO</td>
<td>World health organization</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for disease control and prevention</td>
</tr>
<tr>
<td>FRE</td>
<td>Flesch reading ease</td>
</tr>
<tr>
<td>VADER</td>
<td>valence aware dictionary and sentiment reasoner</td>
</tr>
<tr>
<td>TFIDF</td>
<td>term frequency–inverse document frequency</td>
</tr>
<tr>
<td>RNN</td>
<td>recurrent neural networks</td>
</tr>
<tr>
<td>NLG</td>
<td>natural language generation</td>
</tr>
<tr>
<td>API</td>
<td>application programming interface</td>
</tr>
<tr>
<td>NN</td>
<td>neural network</td>
</tr>
<tr>
<td>LSA</td>
<td>latent semantic analysis</td>
</tr>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
</tbody>
</table>
Abstract

Title: Adaptation of texts to context

Author: Luka Zontar

In this thesis we try to develop a methodology that can adapt texts to target publication types using summarization, natural language generation and paraphrasing. The solution is based on key text characteristics that describe different publication types. To examine types such as social media posts, newspaper articles, research articles and official statements, we use three distinct text evaluation metrics: length, text polarity and readability. While altering key text evaluation metrics, we mostly focus on length due to much research that was done in this field (either with summarization or natural language generation). Using paraphrasing we will try to adjust text readability and polarity that describes reader’s negative or positive orientation towards the topic. The process of text adaptation will be implemented iteratively. The developed methodology will automatize writing articles that are based on existing articles. The more crucial contribution of this thesis is that we help to gain access of harder works to those that cannot understand the origin texts.

Keywords: text adaptation, context-aware, artificial intelligence, text summarization, natural language processing.
Povzetek

Naslov: Prilagoditev besedil kontekstu objave

Avtor: Luka Žontar

V tem diplomskem delu skušamo razviti metodologijo, ki bo znala s pomočjo povzemanja, generiranja naravnega jezika in parafraziranja prilagoditi besedilo ciljnemu kontekstu objave. Za rešitev problema se najprej osredotočimo na ključne lastnosti, ki določajo obravnavane tipe objave. Za obravnavo štirih različnih tipov, ki so objave na družabnih omrežjih, novice, znanstveni članki in uradne izjave, uporabimo tri lastnosti: dolžino, polariteto besedila in berljivost. Pri spreminjanju ključnih lastnosti se osredotočimo predvsem na dolžino, saj se veliko raziskav nanaša na prilagajanje teksta, bodisi s povzemanjem ali generiranjem naravnega jezika. S pomočjo parafraziranja bomo skušali prilagajati berljivost in polariteto besedila, ki priča o negativni oziroma pozitivni naravnanosti bralca k besedilu. Proces prilagajanja besedil bo potekal iterativno. Razvita metodologija bo avtomatizirala proces pisanja člankov, ki temeljijo na obstoječih delih. Še bolj pomembno pa je, da s tem omogočimo dostop zahtevnejših del tistim, ki jih v prvotni obliki ne razumejo.

Ključne besede: prilagajanje besedil, kontekstno-odvisen, umetna inteligenca, procesiranje naravnega jezika.
Razširjeni povzetek

V tem diplomskem delu predlagamo metodologijo za prilagajanje besedil. Pri prilagajanju besedil se najprej osredotočimo na izbrane lastnosti besedila, ki določajo različne tipe objav: dolžina, polariteta (sentiment) besedila in berljivost. Z izbranimi lastnostmi razlikujemo med štirimi različnimi konteksti objav, to so objave na družabnih omrežjih, novice, znanstveni članki in uradne izjave. Medtem ko objave na družabnih omrežjih izstopajo s kratkimi besedili, ki največkrat izražajo le mnenja posameznikov, so raziskovalni članki pravo nasprotje. Med izbranimi tipi objav so najdaljši in naj bi bili objektivni, če izključimo diskusijo in zaključke na koncu člankov. Novice in uradne izjave so si velikokrat zelo podobne, saj so slednje večkrat izhodišče za pisanje novinarskih člankov, pri čemer naj bi veljalo tudi, da so uradne izjave bolj objektivne. Kot testne podatke smo v sklopu tega diplomskega dela izbrali besedila, katerih tematika se nanaša na izbruh virusa COVID-19.

Naša metodologija prilagaja lastnosti besedila s pomočjo povzemanja, generiranja naravnega jezika, zamenjave besed s primernejšimi sinonimi in generiranja parafraz povedi. Na področju prilagajanja dolžine je bilo narejenih že veliko raziskav, zaradi česar sta povzemanje in generiranje besedila ključna procesa pri prilagajanju. Besedilo povzemamo s pomočjo abstraktivnega povzemanja, ki na podlagi vhodnega besedila generira povzetek, pri katerem ni nujno, da vsebuje enake povedi kot izhodiščno besedilo. Povzetek se torej generira na podoben način, kot bi ga generiral človek, t.j. s posploševanjem in izpuščanjem dejstev, ter izpuščanjem in rekonstrukcijo povedi. V kolikor je prvotni članek prekratek, moramo generirati dodatno
besedilo, kar storimo z modeli za generiranje naravnega jezika. V ta namen prilagodimo že naučen model za generiranje naravnega jezika štirim obravnavanim tipom objav. S pomočjo prilagajanja zajamemo lastnosti različnih tipov objav, ki jih s strukturnimi lastnostmi besedila ne bi uspeli (npr. oznake in hiperpovezave na družabnih omrežjih, navajanje virov v raziskovalnih člankih).

Kot pomembni lastnosti poleg dolžine obravnavamo tudi polarliteto in berljivost besedila. Obe metriki hkrati prilagajamo z zamenjavo posameznih besed z njihovimi optimalnimi sinonimi, ki karseda znižajo vsoto absolutnih relativnih razlik obravnavaheh metrik in želenih vrednosti ciljnih tipov objav. Zaradi različnih pomenov, ki jih lahko imajo besede, v tej fazi obstaja verjetnost izgube pomena povedi oziroma besedila, če zamenjava s sinonimom pomeni zmanjšanje vsote absolutnih relativnih razlik polarite in berljivosti besedila. Ker sinonime iščemo po slovarjih, moramo poskrbeti tudi, da bo njihova oblika enaka obliki prvotne besede. Na primer, glagolom moramo spremeniti čas in določiti število. Prvotno obliko besede dobimo s pomočjo orodij v programskem jeziku Python, ki so zmožna dovoljanje besednih vrst. Alternativni pristop k prilagajanju berljivosti je generiranje parafráz in zamenjava povedi z optimalnimi parafrazami, ki zopet maksimalno znižajo absolutno relativno razliko berljivosti in želene vrednosti berljivosti ciljnega tipa objave. V diplomskem delu tudi testiramo, kako zamenjava povedi s parafrazami vpliva na dolžino in polariteteto ter ugotavljamo, da je sprememba teh metrik pri parafraziranju zanemarljiva. Parafraze generiramo z prilagojnim modelom, ki sprejme poved kot vhod in generira seznam parafraz povedi.

Omenjene procese v metodologijo sklenemo z iterativnim pristopom, kjer v vsaki iteraciji izvedemo potrebne omenjene procese. V vsaki iteraciji torej povzemamo besedilo, če je predolgo, oziroma generiramo dodatno besedilo, če je prekratko. Zatem poščemo besede, katerih berljivost in polariteta sta najbolj oddaljeni od želenih vrednosti in jih zamenjamo z optimalnimi sinonimi. Na podoben način prilagodimo berljivost in zamenjamo povedi z optimalnimi parafrazami, pri čemer so povedi za zamenjavo izbrane naključno.
Kot ustavitveni kriterij obravnavamo poleg števila iteracij tudi sprejemljivo napako $\epsilon$. Če je trenutna absolutna relativna razlika do želene vrednosti manjša od določene vrednosti $\epsilon$, izstopimo iz zanke. Število sinonimov in parafraz je vnaprej določeno. Pri objavah na družabnih omrežjih zamenjamo največ 15 besed z njihovimi sinonimi, pri raziskovalnih člankih maksimalno 250, sicer pa 100. Zaradi dolžine objav na družabnih omrežjih pri teh omejimo število parafraz na tri, pri raziskovalnih člankih dovoljimo maksimalno 50 parafraz, sicer pa 20. Zaradi napak, ki jih generirajo modeli pri parafraziranju, povzemanju in generiranju naravnega jezika, po vsaki iteraciji v besedilu uredimo velike začetnice na začetku povedi, dodamo manjkajoča ločila in zamenjamo neustrezno postavljene velike začetnice entitet (kratice in lastna imena).

Zavoljo nepristranskosti zaključkov smo za vsako pretvorbo med štirimi različnimi tipi objav generirali 100 testnih primerov in rezultate povprečili. Pri večini pretvorb med tipi objav zelo uspešno prilagodimo dolžino besedila, ki velikokrat konvergira pod vrednost $\epsilon$. Čeprav v nekaterih primerih ne uspemo prilagoditi polaritete oziroma berljivosti besedila, pri vsakem prehodu uspešno zmanjšamo vsoto absolutnih relativnih razlik omenjenih metrik in želenih vrednosti, kar kaže na uspešnost prilagajanja s pomočjo zamenjave besed z optimalnimi sinonimi. V nalogi tudi predstavimo in interpretiramo začetne odstavke generiranih besedil ter jih primerjamo z izhodiščnimi. Dobljeni rezultati pričajo o uspešnosti metodologije pri prilagajaniu ključnih lastnosti besedila, vendar obenem opozarajemo v dejstvu, da tovrstno prilagajanje ni nujno najuspešnejši pristop pri adaptaciji besedila ciljnemu občinnemu. Model za generiranje naravnega jezika namreč pri generiranju dolgih besedil težko drži rdečo nit, saj zaradi učne množice ne more upoštevati predolgih nizov. Kljub temu se zadovoljimo z dejstvom, da lahko pri generiranih besedilih izpeljemo verigo povezanih tematik po generiranih odstavkih. Poleg tega se razumen človek ne bi lotil prilagajanja objave iz družabnih omrežij v raziskovalni članek, zaradi česar so nesmiselni rezultati neizogibni. Nadaljnja uspešnost metodologije se kaže tudi pri prehodi v ostale tipe ob-
jav, pri katerih bi ponekod lahko brez kakršnikoli sprememb objavili generirano besedilo ciljnemu občinstvu. Razvita metodologija torej zadovoljivo avtomatizira proces pisanja člankov, ki temeljijo na obstoječih delih. Še bolj pomembno pa je, da s tem omogočimo dostop zahtevnejših del tistim, ki jih v prvotni obliki ne razumejo.
Chapter 1

Introduction

1.1 Motivation

With more and more internet usage the amount of textual data on the internet is highly increasing. Every minute the number of posts on Twitter is more than half of a million higher [20]. Well-known newspapers such as The Washington Post publish more than 5000 articles a day [22], while academics publish more than 1.8 million research papers in a single year [7]. Digging through this pile of texts anyone can find something interesting to read. However, since different media target different audiences, an arbitrary article may not be appropriate for everyone. Consequently, already published content is being rewritten and adapted for other target audiences.

Why is targeting audiences so important? When speaking with someone in person, we adjust body language, tone and the words we use, so that the audience understands the message we are trying to send. In a similar manner, we also have to be aware of the target audience when writing. When writing this thesis, we could also consider the target audience all the people that are interested in natural language processing and text adaptation or perhaps just the reviewers that will grade our work. If we know that the reviewers like the statements to be carefully argued and that the text includes lots of technical expressions, we can use that to improve our grade. News-
Adaptation of texts to context, Luka Žontar

Paper articles, for example, target a vast majority of people and should be consequently easily readable. Nowadays, newspapers and social media feed are also forming public opinions about different topics and while social media posts are mostly opinions of individuals, newspapers have to form opinions so that their sponsors gain or at least not lose on popularity. Therefore, newspaper writers have to think of an approach that will satisfy both their main audience and their sponsors as well.

Reid et al. [28] splits audiences into three different categories: the “lay” audience has no expert knowledge and needs background information, the “managerial” can have some knowledge, but still wants background information and the “experts” that are the most demanding and want the writing style to be more technical or specialized. While a particular audience may be fascinated by an arbitrary article, the other groups may feel just the opposite. Therefore, adaptation of texts is crucial if we want to deliver some message or some news to some target audience. A question occurs: “How do we adapt texts to fit the targeted audience?” McMurrey [32] tried to solve the problem of what we should consider when adapting texts to different target audiences. He stated that writer should provide the right information and thus adapt text by omitting unnecessary information, adding examples and generalizing facts. Writers should be careful about the organization of information and write stronger introductions if necessary. McMurrey also emphasizes the importance of writing style. For example, less educated audience tend to be more fond of shorter sentences and less technical terms.

1.2 Goals and contributions

Rewriting and adapting texts to different contexts might be an easy task for experienced writers, but it is still very time-consuming. Furthermore, rookies may find it much harder, since they can struggle in selecting the information that might be relevant to a particular target audience. Nevertheless, adapting text to a specific publication type is highly unlikely to be labeled
as the hardest task for an educated person. A way to deal with words and some common sense should be enough to complete it. However, due to the latter requirement the problem becomes much more complex when trying to automate it.

In this thesis we will try to automatically transform text that is designed for some target audience or some publication type to a text that fits the audience or publication type that we want to target. We will adapt texts by manipulating the key text evaluation metrics from initial values to the mean values of the target publication type that will be calculated from a sample set of articles. We will focus on four different publication types, where each of them, in general, stands out by some metric. Newspapers and social media posts often target a vast majority of people, however, newspaper articles tend to be much longer and probably less opinionated. Research articles stand out by being the longest and much more objective than the previously mentioned publication types. The last publication type that we will consider are official statements. These tend to be more neutrally-oriented, but also harder to read, which is why people usually read the summaries that newspapers publish.

As mentioned before, articles will be adapted by adjusting text evaluation metrics. In this thesis we will focus on three different measures, length, readability and text polarity. Why are these measures important in text adaptation? Too much text can result into reader’s disinterest and if we want to send a message to an audience that might not be interested in the topic, we should just keep it short and sweet. Short texts are often published on social media and consequently we will have to manipulate length in almost every transition from other publication types. As McMurrey suggested [32], omitting unnecessary text is one of the key processes in text adaptation to the targeted audience. In our thesis this will be done by summarization techniques and oppositely if we wanted to increase the level of information, we will generate additional text by NLG methods and finding similar content on Wikipedia and DBpedia. Readability, which estimates the ease with
which a reader can understand a text, and polarity, which estimates whether
the writer is positively or negatively-oriented towards the topic, will be al-
tered using paraphrase generation with neural networks and plain synonym
replacement.

The goal of our thesis will be to develop a methodology that will first
calculate the mean values of the aforementioned text evaluation metrics based
on a set of existing articles of the targeted publication type. Then we will
iteratively alter text metrics, where in each iteration we will try to adapt
initial values so that they converge towards target values. We will allow
some small error $\epsilon$ and limit the number of iterations, so that the method
does not run indefinitely, if the text does not converge towards the desired
values. This methodology should make it easier to adapt texts to different
audiences and thus make more any text available to any type of audience.

1.3 Structure

The thesis consists of six chapters. In the second we start off with examin-
ing different types of publication and calculating textual characteristics by
which publication types differ from one another. The following two chap-
ters are dedicated for the three different methods that will be used for text
adaptation. In Chapter 3 we start with text summarization and generation
of additional text that has to coincide with the existing. The next method
described in Chapter 4, tries to adjust polarity and readability of the text to
the target audience by paraphrasing the sentences in text. In Chapter 5 we
put it all together by developing a methodology which adapts texts combing
the previously mentioned methods and the calculated metrics that define
different publication types. In Chapter 6 we present conclusions and ideas
for further work.
To really understand the problem that we will be discussing, we first have to go through existing related work. In this chapter we will look at some of the most important existing articles that helped us complete this thesis. As mentioned in the introduction, our methodology will try to adapt text using text evaluation metrics. Without knowing, which metrics we want to adapt, we cannot discuss the methods that we will use to do that. Therefore, we will first look at researches about existing text evaluation metrics and which text characteristics they describe. Next, we will focus on length manipulation by looking at research about summarization and natural language generation. To conclude, we will name and discuss approaches to automatic paraphrase generation.

2.1 Text evaluation metrics

2.1.1 Readability and corpora

Štajner et al. [29] calculated measures of readability for different corpora (e.g. news, health, fiction) and found that all indices score similarly for similar topics. Furthermore, they investigated the correlation between these readability measures and found that they are linearly correlated to each other. Lastly, in that article a test was conducted, where researchers measured cor-
relation between readability measures separately for each of the corpora and concluded that correlations are not a consequence of genre dependence.

2.1.2 Structure-based metrics

Kiefer in her article [14] describes a few key characteristics that help determining the quality of our text considering its structure. Many structural metrics that Kiefer describes alter text readability. For example, percentage of abbreviations and uppercased words require more prior knowledge from the reader. Another structural characteristic is lexical diversity, which gives us an idea about how rich the writer’s vocabulary is. On the other hand, using lots of synonyms makes a text less readable and more complex. Interestingly Kiefer discovered that Twitter posts are usually very lexically diverse whereas more complex texts tend to use smaller subset of vocabulary [14].

2.1.3 Readability metrics

Some texts are easy enough for children to read them, however other texts are written for a very narrow group of people with a very specific education and are thus hard to understand for the vast majority of people. Andova in her bachelor’s thesis describes readability measures that are based on statistics and previously mentioned text structure [2]. In our thesis, structure-based readability measures will be accurate enough. Nevertheless, lots of other academics are more fond of looking at readability calculation as a classification problem. Martinc et al. [21] describe both supervised and unsupervised neural approaches to calculate text readability.

Unsupervised neural approach

Martinc et al. [21] proposed a peculiar methodology to calculate readability using an unsupervised neural approach. Firstly, we have to split a text to sentences. After that we evaluate each sentence in the following manner:
1. Calculate word negative log-likelihood (WNLL) that gives us a readability estimate for each word:

\[ WNLL = -(y_t \log y_p + (1 - y_t) \log(1 - y_p)) , \]

where \( y_p \) denotes predicted probability by the language model and \( y_t \) denotes the actual probability.

2. Sort the words according to WNLL score in ascending order

3. Calculate ranked sentence readability score (RSRS)

\[ RSRS = \sum_{i=1}^{S} \frac{\sqrt{i} \cdot WNLL(i)}{S} , \]

where \( S \) denotes sentence length and \( i \) represents the rank of word.

Authors state that the main idea of RSRS score is to avoid reductionism of traditional readability formulas by including discourse cohesion and background knowledge through language-model based statistics. They assumed that low discourse cohesion has a negative effect on the performance of the language model. Experiments resulted that some corpora had better results using RSRS, while for other corpora results were better when using traditional readability metrics.

**Supervised neural approach**

In their article Martinc et al. [21] also tested three different supervised neural approaches:

- Bidirectional Long short-term memory network (BiLSTM) - using RNN approach for classification and bidirectional LSTM layer (LSTM layers that read text in opposite directions).
- Hierarchical attention networks (HAN) - using two level attention mechanism while taking into an account hierarchical structure of the text.
- Transfer learning - using pretrained BERT transformer with 12 layers of size 768 and 12 self-attention heads. A linear classification head was
added on top of the pretrained language model and the classification model tuned for 3 epochs for each data set.

Once again results vary a lot in different corpora. While researchers listed BERT as the most accurate classifier achieving the accuracy of 83.93% in one corpus, it performs the worst in other corpora. Authors claim that semantic similarity in other corpora is the cause of such variance.

2.1.4 Sentiment analysis

Until now we were focusing on facts and ignoring opinions. Writer’s subjectivity towards the topic can be expressed with the usage of more positive or negative words. Subjectivity has to be considered when adapting texts to fit the targeted audiences. Even though some experts claim that no article can be written without being subjective, it is still important to know how subjective someone is. Obviously every writer decides by himself what he will put in the article. We will be more interested in how writers include their opinions in different publication types. Sentiment analysis denotes whether the writer has positive or negative affections towards the facts that are described in the text. Feldman [8] discusses several approaches of sentiment analysis based on the basic unit that we will be classifying. We can classify whole documents, sentences or even particular aspects. Furthermore, we have to consider comparative statements such as less and more. As the most crucial field of sentiment analysis Feldman states sentiment lexicon acquisition, where words can be expanded in a set of words using corpus-based approaches or WordNet.

2.1.5 Length

Length is one of the key text evaluation metrics, because a lot of research has been done to alter it. Moreover, it is the most visible factor of the aforementioned publication types. For example, Belder and Moens calculated the average sentence length of 100 Wikipedia articles, which resulted in 21.6
words per sentence [3] and furthermore found that longer sentences usually correlate with more complex texts.

2.2 Length manipulation

2.2.1 Extractive summarization

According to Allahyari et al. [1] there are several different approaches in extraction based summarization. The simplest form of extractive summarization aim to identify the words that describe the topic of our text. We can do that with several techniques such as finding the topic signature with log-likelihood ratio or TFIDF method that calculates word probability inside a document. This topic signature can then be used to find the most important sentences.

Another approach is Latent Semantic Analysis (LSA) that works with term-sentence matrices and evaluates word weights for each term. Then we apply singular value decomposition to this matrix, where matrix $U$ represents a term-topic matrix, matrix $\Sigma$ has weight of topics on diagonal and matrix $D$ describes how much a sentence represents a topic. LSA method splits the text to topics where the summarized text only keeps one sentence per topic.

Summarization can also be interpreted as a classification problem, where sentences are classified as summary or non-summary sentences. Another way is to create a graph, where sentences are vertices and edges between those vertices hold the information about similarity. If we only connect vertices of which similarity is higher than some threshold value, we can simply extract the sentences with the most edges as our summary.

2.2.2 Abstractive summarization

Based on recent works done in this field, we can divide abstractive summarization into two different approaches. On one side, there are abstractive approaches that try to summarize the text by, firstly, structuring the text into
some logical form such as graphs, trees and ontologies. Graph-based methods can generate summaries that are more informative and better formed. For example, Opsinosis [10] iteratively searches through subgraphs and tries to find a valid sentence. Valid sentences are then evaluated through redundancy where the most repetitive sentences are excluded from the text. Moreover, Huong Le and Tien Manh Le published a promising research [17] where they presented a new abstractive summarization technique. They used discourse rules, syntactic constraints and a word graph to summarize texts.

The other group of abstractive approaches focuses on text sentiment. Semantic graphs (e.g. rich semantic graph [23]) can be used to generate a summary. In such graphs nodes represent verbs and nouns and the links between these represent the semantic relationship. After the graph is generated, it is reduced to a more concise graph. The reduced graph is used to generate final sentences to be used in the summary. Text-to-text generation [11] is another summarization technique that is based on information item, which is the smallest element of coherent information in a text. It generates subject-verb-object triplets using syntactical analysis of the text. The sentences are then generated with a language generator and either excluded from the text or not, which is decided by document frequency score.

2.2.3 Natural language generation

Chen et al. [4] explain how we can train network based sequential inference models that use chains of LSTMs. Recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) have proven to be a very efficient tool when coping with language generation. Using subsequences of texts they try to determine which character should follow after inputting a sequence.

Radford and his colleagues [25] used unsupervised multitask learners for model learning and developed models that provided state of the art results in multiple language modeling datasets. The original model was trained to predict the next word based on 40GB of Internet content. They concluded that large training datasets and models trained to maximize the likelihood
of a sufficiently varied corpus can learn a surprising amount of tasks, while no supervision is needed in training.

2.3 Paraphrase generation

2.3.1 Automatic paraphrase generation

Because we cannot reasonably discretize the meaning of a sentence, paraphrasing is a hard task for a computer. Nevertheless, Prakash et al. [24] suggested a method to generate paraphrases using a stacked residual LSTM network. Evaluation results demonstrate that their model outperforms sequence-to-sequence, attention-based and bi-directional LSTM models. On the other hand, sequence-to-sequence models with some optimizations are also very efficient. Egonmwan and Chali [6] proposed a framework, where they combined transformer and sequence-to-sequence models. While sequence-to-sequence models are generally more efficient than synonym replacement approaches, transformer layers help to catch long-term dependencies and thus improve its results.

Another approach was proposed by Goutham [27], who used a pre-trained text-to-text transfer transformer to generate paraphrases of questions. Text-to-text transformer can also be used in paraphrase generation and it simply accepts an input text and returns a modified input text. How does the model know what kind of output it should return? The model is learned with a dataset of a sample input and the expected output. As we trained the model for summarization with a dataset of initial texts and its summaries, Goutham did the same for paraphrasing. He used a dataset of questions from Quora that were labeled as duplicates and used those for model learning. The method proved to be very efficient by generating state of the art results.
Chapter 3

Text evaluation metrics

Because we give meaning to the text we write, natural language processing is a fairly hard problem that many researchers are tackling. Even though the artificial intelligence has come a long way in the past few years, we seem to be a long way from finding a method to store the meaning of a text so that an artificial intelligent system can understand it as people do. Until then we will have to make use of simpler methods to evaluate text and not corrupting the content while adapting it. To adapt our text to the target audience, we will have to define our problem numerically. We will do that by choosing some key text evaluation metrics with which we can distinguish different target audiences. This will be crucial in this thesis, because the calculated mean values will be used to evaluate the efficiency of the developed methodology in the end.

3.1 Structure

In this chapter we will present which publication types we will be examining in our thesis and try to determine how they correspond to different target audiences. This way we will find out how can different audiences be associated with the chosen text evaluation metrics. Then we shall calculate the mean values of chosen text evaluation metrics, graphically show results and
interpret them.

3.2 Publication types

In this section we will present the publication types that we will examine in this thesis. Data used for mean value calculation holds text that contain COVID-19 related content. This will not only simplify collecting appropriate datasets, but will also minimize the effect of variables that we will not take into account in text adaption.

3.2.1 Social media

The first source of textual data that we will consider in this thesis is social media. Nowadays a vast majority of people use social media to publish their opinions on everyday matters, which is why we expect the feed from social media to be more readable and also more opinionated. Due to the fact that most social media providers limit the number of characters that users can publish in a post, we also presume that texts will be shorter in comparison with other publication types.

From dataset that will be used to calculate mean values of text evaluation metrics we extracted 150 rows, where each row represents text from an independent social media post from Twitter with COVID-19 content [16].

3.2.2 Newspaper articles

Online newspaper articles should be more readable, because they target the general public. Even though newspapers were originally meant to only talk about facts that occur around the world, more and more reporters are obviously expressing their opinions in their articles, which is why we should expect a lot of variance in polarity. Considering length, newspapers should be longer in comparison to social media posts.
From dataset that will be used to calculate mean values of text evaluation metrics we extracted 150 rows, where each row represents text from a newspaper article with COVID-19 content [13].

### 3.2.3 Research articles

Research articles are probably a bit less opinionated, because there is no room for opinions. Scientists should only state facts, except in discussion and results interpretation, where they can include some personal insights. Researchers also like to show off with their rich vocabulary and reduce readability by adding lots of synonyms. Research articles tend to be much longer than newspapers articles.

From dataset that will be used to calculate mean values of text evaluation metrics we extracted 150 rows, where each row represents text from research article (just body of the article) with COVID-19 content [31].

### 3.2.4 Official statements

Official statements are those texts that governmental authorities release to the press, when they want to send a message to their citizens. While the rest of publication types tend to be quite opinionated, official statements are bound to not express feelings towards the topic, which is why we expect polarity to be somewhere near zero with little variance between different texts.

From dataset that will be used to calculate mean values of text evaluation metrics we extracted 150 rows, where each row represents text from an official statement of either WHO or CDC organization with COVID-19 content [19].

### 3.3 Used metrics

What kind of text characteristics can be used for adapting texts? For example, percentage of uppercased words can be an identifier of text quality.
However, we should ask ourselves, if we can adapt it, so that it corresponds with the target publication type. Even if we would try to adjust the number of uppercased words the average texts of targeted publication type, how do you generate extra uppercased words meaningfully? Moreover there are some characteristics, such as lexical diversity, that can be harder to adjust due to lack of repeated words or synonyms. In our thesis we will use three different text characteristics that will help us differentiate between publication types.

3.3.1 Length

Length of the text is a measure that is almost unique for different publication types and can also be manipulated with summarization and natural language generation. Since some of the methods of text generation are based on tokens, length will be calculated as the number of words that our text holds. For this we will use Natural Language ToolKit, which contains a function that transforms a string into a list of tokens using regular expressions.

3.3.2 Readability

Readability is a bit more difficult characteristic to adapt because we would have to deliberately worsen the text if we would want to decrease readability. It is also hard to make a computer redefine a sentence in such a way that it becomes easier to read. Nonetheless, we will include it into our thesis. As mentioned before there exist lots of different scores to evaluate text readability using text statistics. We will use Flesch Reading Ease score [9]. It can be calculated by the following formula:

$$\text{FRE} = 206.835 - 1.015 \left( \frac{\text{totalWords}}{\text{totalSentences}} \right) - 84.6 \left( \frac{\text{totalSyllables}}{\text{totalWords}} \right)$$

Flesch Reading Ease score [9] is a readability score, where higher scores indicate material that is easier to read and lower scores indicate material that is harder to read. The formula that is used to calculate tells us that
the score is dependent on the number of words, the number of syllables and the number of sentences. Furthermore, we can see from the formula that the score is decreased by sentences that contain more words and words that contain more syllables. We can all agree that shorter sentences tend to be more understandable and that the longest word in English language “pneumonoultramicroscopicsilicovolcanoconiosis” that contains 19 syllables makes this sentence and thus this thesis much less readable.

3.3.3 Polarity

While subjectivity tries to determine the degree to which writer expresses his opinion towards the topic, polarity score determines whether writer’s opinion about it is positive or negative. In our thesis we will be using VADER (Valence Aware Dictionary and sEntiment Reasoner), where the compound score is computed by summing the valence scores of each word in lexicon and then normalized to the interval [-1, 1]. Since polarity can be easier noticed in context, we will calculate polarity by evaluating each sentence and then calculating the mean value of all the sentences.

3.4 Findings and interpretation

In this section we will present the initial values of used text evaluation metrics for the dataset of each publication type and try to interpret them. In the subsequent chapters, we will use calculated text metrics to iteratively adjust text to the target audience. The initial values are shown in bar charts in Figures 3.1-3.3.

Considering length, we correctly assumed that official statements and news articles will be longer than social media posts. We also expected research articles to be much longer than other types of texts. On barcharts from Figure 3.1 we can also notice that research articles have a very high standard deviation, which means that research articles are not necessarily as long as the average score suggests, but can also be longer.
As can be seen in Figure 3.2, official statements and research articles are harder to read, whereas news and social media posts that mostly target the general audience are much easier to read. Even though social media posts are easier to read, they still require 8th or 9th grade education. Newspaper articles are even harder since they require some college education to completely understand them. Research articles and official statements are understood by college graduates based on Flesch Reading Ease score.

From Figure 3.3 we can notice that social media posts have surprisingly scored the lowest in polarity along with research articles. A very high standard deviation in social media posts dataset tells us that social media posts are in fact full of negatively and positively-oriented texts, but somehow different text polarities almost cancel out. We could explain that there are two groups of tweets: those that speak badly about government’s reactions and
Figure 3.2: The initial value of readability score for different publication types

surviving quarantine, and those that are complimenting healthcare workers, trying to stay positive etc. The sizes of those groups are almost the same. In newspaper articles dataset positivity prevails, but not excessively. Curiously enough official statements score quite high when it comes to polarity. Why did WHO and CDC use so many more positively-oriented phrases? Perhaps they simply tried to encourage people and not spread panic.
Figure 3.3: The initial value of polarity score for different publication types
Chapter 4

Length manipulation

As described in Chapter 3 we will use length as one of the text evaluation metrics in text adaptation to context. Since our goal will be to alter length of an existing text without changing its meaning, we will have to either summarize or extend the initial text. In this chapter we will go through some of the methods that will help us adapt the initial text to the target audience by either shortening or extending the existing text and thus altering length as one of the key text evaluation measures.

Therefore, the ability of automatic shortening of texts without losing key information will play a very important role in the development of our methodology. We can classify summarization as a complex task due to inability to discretize the meaning of text. Therefore it is hard to determine which words are crucial to stay in the text and which are not. Fortunately, lots of successful research was made in the area of summary generation. Experts [1] divided text summarization into extractive and abstractive approaches, where the first extracts the most significant sentences from the text and the latter tries to summarize text in a new (more human-like) manner. In our thesis we will describe the pros and cons of each approach in text adaptation to context.

On the other hand, we cannot imagine social media post to be additionally summarized into a research article. Therefore, we will also have to generate
new content related to the existing, which seems as a much harder problem for a computer in comparison to summarization. Intelligent systems have proven to be very useful in generating new textual data using natural language generation techniques. In this chapter we will fine-tune a pre-trained model on each of the aforementioned publication types in order to get better results. However, such algorithms are not the only way to generate extra content. For comparison, we will also develop a method that finds the keywords of a text and searches for similar articles on Wikipedia and DBpedia using generated keywords and thus generates additional content by finding similar articles.

4.1 Structure

Starting this chapter, we will first go through an extractive summarization approach. We will explain the benefits and drawbacks of such approach. Then we will compare the mentioned approach with an abstractive approach of summarization. In the following, we will focus on text generation starting with natural language generation techniques, where we will explain the concept and developed models. To conclude the chapter, we will try to generate text by finding articles from Wikipedia and DBpedia using keywords-based search.

4.2 Extractive summarization

Using extraction summarization techniques we will be able to shorten the text, if the initial value of the targeted publication type that we calculated in Chapter 3 is smaller than the current length of the input text. Allahyari et al. [1] explain that extractive summarization generates summaries by extracting sentences from the original text. Text can only be shortened by leaving out whole sentences. Writing summaries usually includes changing sentences not only by replacing words with synonyms but also by generalization and omission of facts to make a text as concise as possible. However, despite
the recent rise of neural networks in natural language processing and generation, these methods still provide better results than abstractive approaches in some cases.

4.2.1 Used methods

In this section we will show how extraction based methods work by going through the used method and providing sample results.

Firstly, we transform text into a list of sentences, where each sentence is represented as a list of words excluding stop words. To rank similarity scores we will use cumulative frequency of the words in the sentences. Ranked sentences are then sorted in the descending order. We extract the first (most important) $n$ sentences as our summary [12].

The example in Figure 4.1 shows a generated summary that holds 105 words from a text that originally contained 183 words. The first and the last sentence could be considered as the most important as they hold many key concepts. On the other hand, we probably would not include the second sentence in the summary, because it does not express anything important. Moreover, when traversing from the first to the second sentence, a big difference between key messages of the sentences can be noticed. To conclude, sentence extraction works well when trying to get the sentences that hold the most information, but the method has problems producing summaries that hold a common thread throughout the whole text. If the sentences are very related between each other, the generated summaries tend to be of a high quality. For instance, if the first three sentences in the example in Figure 4.1 were talking about the video as well, the summary would be much more satisfactory.
A 26 July report from the White House coronavirus task force identified 21 states as being in the "red zone", meaning that they had reported more than 100 new cases per 100,000 residents in the past week.

Twenty-eight states are in the "yellow zone". Vermont is the only state in the green, with fewer than 10 cases per 100,000 people in the past week.

The report calls for further closures by state governors to contain the spread of the outbreak.

The video retweeted by Mr Trump showed doctors speaking outside the US Supreme Court building at an event organised by Tea Party Patriots Action, a group that has helped fund a pro-Trump political action committee.

In the video, Dr Stella Immanuel, a doctor from Houston, says she has successfully treated 350 coronavirus patients "and counting" with hydroxychloroquine.

The president said on Tuesday: "I think they're very respected doctors. There was a woman who was spectacular in her statements about it."

A 26 July report from the White House coronavirus task force identified 21 states as being in the "red zone", meaning that they had reported more than 100 new cases per 100,000 residents in the past week.

The video retweeted by Mr Trump showed doctors speaking outside the US Supreme Court building at an event organised by Tea Party Patriots Action, a group that has helped fund a pro-Trump political action committee.

In the video, Dr Stella Immanuel, a doctor from Houston, says she has successfully treated 350 coronavirus patients "and counting" with hydroxychloroquine.

Figure 4.1: Extractive summarization example
4.3 Abstractive summarization

As mentioned in the introduction of Chapter 4, an alternative to extraction based summarization is abstractive summarization. We will use it for the same function, i.e., to shorten the input text to the initial length of the targeted publication type. In contrast to extractive summarization, this approach tries to generate summaries in a brand new way, much more like people do. Abstractive summarization techniques are based on deep-learning based models that are used in natural language processing as well. In this section we will first explain our approach and then interpret a few sample results.

4.3.1 Used methods

We used the T5 text-to-text model [26]. It’s an encoder-decoder model that uses transfer learning on a model that is firstly pre-trained on a data-rich task and then fine-tuned on a downstream task. T5 text-to-text accepts the following parameters:

1. number of beams for beam search (heuristic search algorithm that optimizes graph search): 4
2. no-repeat ngrams (ngrams that can only occur once): 2
3. min length: 90
4. max length: 110
5. early stopping (beam search stopped early when at least num_beams sentences are finished): False
6. repetition penalty (if a sequence repeats): 2.5
7. length penalty (if we are out of min-max boundaries): 1.5

We generated a summary of 92 tokens from the original text that contained 183 tokens. As it can be seen in Figure 4.2, the model mostly uses the content from the original text. In the first sentence of the summary we can see that “identified” was transformed to “says”, which is acceptable in this
A 26 July report from the White House coronavirus task force identified 21 states as being in the "red zone", meaning that they had reported more than 100 new cases per 100,000 residents in the past week. Twenty-eight states are in the "yellow zone". Vermont is the only state in the green, with fewer than 10 cases per 100,000 people in the past week. The report calls for further closures by state governors to contain the spread of the outbreak.

The video retweeted by Mr Trump showed doctors speaking outside the US Supreme Court building at an event organised by Tea Party Patriots Action, a group that has helped fund a pro-Trump political action committee.

In the video, Dr Stella Immanuel, a doctor from Houston, says she has successfully treated 350 coronavirus patients "and counting" with hydroxychloroquine. The president said on Tuesday, "I think they're very respected doctors. There was a woman who was spectacular in her statements about it."

Figure 4.2: Abstractive summarization example using T5 text-to-text model case. The first sentence is sensibly merged from two different sentences from the original text. Some sentences are obviously left out. The only sentence that we miss in the summary would be: "Dr. Stella Immanuel says she has successfully treated 350 coronavirus patients "and counting" with hydroxychloroquine". Since this sentence is missing, the last sentence in the summary does not really make sense. Nevertheless the use of "President Trump:" instead of "The president said on Tuesday:" looks very encouraging. Overall, the summary holds most of the key concepts and is also well-formed.
4.4 Natural language generation

The opposite operation of text shortening is generating additional text. We will need to generate new text in order to adapt length of input texts that are shorter than the initial length of target publication type. One way of doing that is by using natural language processing and generation techniques. Using neural networks we can teach a model by transforming natural language to structured data and then we can use that model to generate additional text based on some input string, which is first encoded into structured data. The model then generates the most probable sequence based on the training dataset which can be transformed into natural language. In this section we will see how natural language generation methods can be used in text adaptation to context and whether or not the generated content is meaningful.

4.4.1 Used methods

In this thesis we will fine-tune a pre-trained NLG model that is based on the aforementioned unsupervised learning. While the model would work sufficiently if we just wanted to generate text, we have to fine-tune it, because we want the generated text to have text evaluation metrics similar to the initial values of target publication type. This is why we fine-tuned the model with four different datasets, where each held sample texts of a different publication type. For inferencing, we have to first encode the input sequence using a tokenizer, which the model will consider when generating additional text (i.e., the output should be similar). Then we generate a list of possible next words using the model and narrow down the list using top k sampling, which sorts the list by probability and sets zeroes to the elements below k-th token. The chosen word maximizes the likelihood of the final text. We continue generate as many words as needed and finally decode the output sequence using the tokenizer.

Once the model is fine-tuned, we can generate an arbitrary number of words based on the sequence we input. In the following, we will try and
generate sequences of 100 words using different models, where the initial sequence will be “The Coronavirus outbreak is a big problem in the United States.” In Figures 4.3-4.6 the text colored with red is the initial sequence, the rest is generated using NLG models.

Model fine-tuned on social media posts

The Coronavirus outbreak is a big problem in the United States.
A new survey shows that nearly 1,200 people have tested positive for #COVID19 so far and more than 2 million Americans are at risk of becoming infected! We must work together to contain this spread & protect our families from...
https://t.co/zBXk3Qm6lyMCoronavirus @NHS England confirmed 713 cases on Wednesday afternoon (26th July)

Figure 4.3: Generated text using model fine-tuned on social media posts

When writing a tweet we cannot pass writing emojis and hashtags, mentioning others and adding links in between the text. In Figure 4.3 we can see that the generated text includes hashtag “#COVID19”, mentions NHS England, which seems to be a healthcare service provider and it also contains a link to an unknown website. Furthermore, the text generated as the continuation of the input sequence is very well related to the context. Overall it looks like a social media post that we could publish. It would be better if the sentence “We must work together to contain this spread & protect our families from...” ended with something meaningful instead of the ellipsis and if the generated link-like text would actually take us to an existing website, but the results seem promising nevertheless.

Model fine-tuned on research articles

One of the first things that come to mind when writing an article is bibliography. In the example from Figure 4.4 the model recognized that the texts from the input dataset of research articles contained lots of citations and included some in the generated text. Once again, the generated text seems to
The Coronavirus outbreak is a big problem in the United States. However, since we still don't have all the data and techniques to address this epidemic adequately (e-mailing CDC with questions or suggestions), it's essential that you know where your local health department can help prevent further spread of this virus: Infectious Diseases Society for Prevention [1], Centers For Disease Control & Protection [2], National Center on Chronic Infections and Respiratory diseases [3],[4].

Figure 4.4: Generated text using model fine-tuned on research articles

be well related with the context. How could we include generated text in an actual research article? We could imagine the text being in the introduction of an article that tries to find out the means of communicating Coronavirus related content.

Model fine-tuned on official statements

The Coronavirus outbreak is a big problem in the United States. It's been spreading to people across all 50 states and more than 25 countries, including Mexico, where it has killed 16 patients so far this year. Over 200 US residents are under quarantine for their home country of Japan after coming into contact with three cases that were imported from there last week – one person who tested positive at Japanese hospitals before returning back north; another infected by an individual while on vacation here during his stay.

Figure 4.5: Generated text using model fine-tuned on official statements

Official statements from governmental bodies tend to talk about the situation of their own citizens and issues that impact them. In our example that is shown in Figure 4.5, we can imagine the text being published by the United States government, because it talks about the situation in America, their citizens and their neighbor Mexico. Once again the generated text is meaningful and all the content relates to one topic: Coronavirus and the USA.
Figure 4.6: Generated text using model fine-tuned on newspaper articles

Model fine-tuned on newspaper articles

The text from Figure 4.6 looks like a common Coronavirus report from the beginning of the outbreak. It talks about the current situation, mentions the online communities, which we certainly would not see in research articles or official statements and lastly it even mentions the increased usage of “#COVIDFree" hashtag. In social media posts we would probably not talk about how hashtags are being used more, but rather just use them ourselves.

Overall, the natural language generation method not only generated meaningful data, but also caught some obvious characteristics of different media that would be hard to include otherwise.

4.5 Text generation with keyword-based search

As mentioned before, if the target audience expects more content, we have to deliver it and thus generate additional data if the input text is shorter than expected for target publication type. Similarly to natural language generation, we will try to generate additional text in this section as well. However, in this section we will not use AI methods to do that. When adapting research articles it is also very important to check out the articles that cite or are cited by this article, because they usually hold similar content [1]. We will use a similar approach, but we will search for additional data on Wikipedia and DBpedia using keywords of the initial text. The concept
is commonly used in extractive summarization, but we will try to use it for additional text generation.

Firstly, we generate keywords that represent the main concepts of our text. We will use a method that makes use of TextRank to extract keywords [18]:

1. split text to tokens,
2. filter stopwords,
3. exclude words that are not nouns or verbs (based on part-of-speech tags),
4. create n-sized windows where each word in a window has an undirected edge to the rest of the words in the window,
5. calculate weight for each word based on the graph that we create from windows,
6. extract words with highest weights as keywords.

Once we calculate keywords we only have to search for additional data using Wikimedia API (https://www.mediawiki.org/wiki/API:Main_page/en) and DBpedia API (https://wiki.dbpedia.org/rest-api).

We will try to generate additional text from the original text from the example in Section 4.1. Firstly, we extract 10 keywords and corresponding weights.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronavirus</td>
<td>1.82</td>
</tr>
<tr>
<td>state</td>
<td>1.64</td>
</tr>
<tr>
<td>doctors</td>
<td>1.62</td>
</tr>
<tr>
<td>cases</td>
<td>1.52</td>
</tr>
<tr>
<td>action</td>
<td>1.34</td>
</tr>
<tr>
<td>report</td>
<td>1.30</td>
</tr>
<tr>
<td>video</td>
<td>1.23</td>
</tr>
<tr>
<td>immanuel</td>
<td>1.10</td>
</tr>
<tr>
<td>doctor</td>
<td>1.10</td>
</tr>
<tr>
<td>party</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Table 4.1: Keywords from texts sorted by descending weight
If we take for example “coronavirus” keyword, we gain articles from Wikipedia such as: Coronavirus, Coronavirus disease 2019, Coronavirus party etc. From DBpedia we gain the definition of the word, data about bird diseases, Pneumonia, severe acute respiratory syndrome, animal virology etc. We do not want to include that much textual data and moreover, some articles do not have as much in common with our text as the others. To decide which sentences will be input in our text, we calculate similarities between all the generated texts and the original text and include only those that share the same context.
Chapter 5

Paraphrase generation

In addition to length we predicted that readability and polarity are just as important when it comes to text adaptation to target audience. However, since we will be using the methods described in Chapter 4 to manipulate length, the most straightforward way to manipulate readability and polarity without altering length is to switch words with appropriate synonyms or change the structure of the sentences. In this chapter we will be using paraphrasing and synonym replacement to adapt text polarity and readability to the targeted values. Paraphrase is a rewording of something written, where we can either change its structure or just replace the words with their synonyms. Since English grammatical rules are complex to encode into an algorithm, it is also very hard for a computer to efficiently alter text structure. Firstly, we will try an approach that replaces words with their synonyms found in the dictionary and this way generate a text of which text readability and polarity scores are closer to the targeted values. The second approach uses neural networks to generate a paraphrase from a sentence.

5.1 Structure

In this chapter we will first see how we should pre-process our data to get better results. Next, we continue with development of an approach that
merely replaces the words with optimal synonyms. In the following, we will try out the proposed approach by altering readability and polarity, each metric in a separate section. Finally, we will fine-tune a pre-trained model for paraphrase generation using neural networks and interpret a few sample results.

5.2 Preprocessing

Using artificial intelligence pre-processing is crucial to get the expected results. In the following our methods will use the following pre-processing:

1. add spaces between puctuations and the words next to them,
2. remove extra spaces,
3. remove characters that are non-alphanumerical or punctuations.

We also have to add one more step in synonym replacement approach. Because we are searching for synonyms through the dictionary, we have to exclude words that usually do not have synonyms. Thus we will exclude numbers, stop words and named entities.

5.3 Synonym replacement

In synonym replacement approach we first generate a list of tokens based on a pre-processed text. The list is then sorted in a descending order using a defined evaluation metric (in our case it will be either readability or polarity). Then we walk through the ordered list and with each iteration we find a set of synonyms of a word. Once again we go through the set and find the optimal synonym. What is the optimal synonym? If we replace the original word with the optimal synonym, we generate a text of which evaluation metrics are the closest to the target evaluation metrics. If there is no synonym that improves our situation, we continue with the next word.
Another detail that we have not taken into account yet is the form of the word. We have to be careful to conjugate the verb into a correct form, consider whether nouns are in singular or plural form and whether an adjective is superlative. Luckily for us, libraries that can help transform a word into its original form exist. To identify the initial form of the word we classify Part-of-speech tag \[15\]. These tags hold the information about the type of the word (e.g. verb, noun, adjective) as well as the form (e.g. form of the verb such as present perfect and whether the noun is plural or singular).

5.3.1 Readability manipulation

We will try the synonym replacement approach and adjust readability of a text. As mentioned before, some publication types target wider audience. All the people that find themselves in that audience have to be able to understand the text with no problems. Long sentences and words with lots of syllables are indicators of a hard-to-read text. However, how can we automatically alter readability in text? We will try to do it by replacing words with more appropriate synonyms.

We first generate a list of tokens from the pre-processed text. The list is then sorted in descending order by the number of syllables of words. Flesch Reading Ease score evaluates readability based on total number of words, total number of sentences and total number of syllables. Length manipulation already considers the number of words and sentences, which is why we have to take care of the number of syllables in this section. If we wanted to increase readability and thus make the text easier to read, we would have to replace words with synonyms that have less syllables. If we were adapting text to a publication type that usually has lower readability, those synonyms would have to hold more syllables. In case that the readability score would have to be changed by a significant amount, we could also use length manipulation techniques that were explained in Chapter 4.

As an example we will apply the method on the first paragraph of The Velveteen Rabbit, written by Margery Williams in 1922. The initial read-
ability score is 69.76, which means that 8th or 9th grader can read it. It signifies a plain, easily understood English. We will try to make the text less readable, where the target readability score will be set at 30 (that is, the reader requires college education to read the text).

Figure 5.1: Synonym replacement of readability manipulation example

In Figure 5.1 text readability is altered in a sample text using synonym replacement approach. While we successfully drop readability to the final score of 30.37, we also replace words with inappropriate synonyms, because the meaning of some words is changed. In the following we will comment some replacements:

- coat → laboratory coat - the text is clearly not speaking about laboratory coats.
- paws → Calyptridium umbellatums - the synonym represents a flower while the original word indicates a part of the body.
- sprig → ornamentation - ornamentations are used for decoration and while sprigs can be used for that as well, in our example rabbit holds a twig or a small branch of some plant.

While some synonyms lightly change the meaning of text, the message is still understandable. We should be aware that for more efficient synonym replacement, we have to take into consideration content and not just replace single words. Moreover, replacing single words to adjust readability is most likely not the best option. While considering content is one of the crucial
improvements, we should also look at the text as a whole, which would make the adaptation much more complex.

### 5.3.2 Polarity manipulation

Similarly to the previous subsection we will try to manipulate polarity with this synonym replacement approach. Polarity is part of the sentiment of the text and because computer has limited ability to understand the sentiment let alone alter it meaningfully, we will first use a similar synonym replacement algorithm as in Subsection 5.3.1. The list is ordered by word polarity. Even though word polarity is our best option for sorting, we have to be aware that it is not the perfect metric. For example, the word “good” usually indicates something positive, however when put together as “good morning”, the word seems much more neutral. In other words, polarity score of a single word is less reliable due to the multiple meanings that a word can have. Then we try to replace words with synonyms and evaluate text polarity of the transformed text. If it is closer to the target polarity, we apply the change.

Once again we will apply the method on the first paragraph of The Velveteen Rabbit. The initial polarity score is 0.43 and we will try to make it more negative. The target polarity score will be set at 0.1.

![Synonym replacement of polarity manipulation example](image)

Figure 5.2: Synonym replacement of polarity manipulation example

The results are much better in this case. We can see that synonyms are input reasonably, i.e., no synonym that was switched changes the meaning of the text. While we cannot really see the negativity that the majority of
replacements contributed (e.g. we cannot notice the difference in polarity for words “really” and “genuinely”), we notice that the word “fat” was replaced by the word “gross”. The replacement holds much more dismissive meaning. In Cambridge Dictionary [30] “gross” is defined as: “extremely fat or large and ugly”, which justifies the substitution. The process went through all the words and their synonyms and minimized polarity to 0.16. If we include some error $\epsilon$ (e.g. $\epsilon = 0.1$), then we successfully adapted text polarity.

5.4 Paraphrase generation using fine-tuned text-to-text transformer

An alternative approach to synonym replacement would be to consider a sentence as a whole, meaning that we generate paraphrases of a sentence and replace the original sentence with the most appropriate paraphrase. This way we should be able to alter both readability and polarity scores. In this section we will look at some practical cases of the implemented method.

5.4.1 Used methods

This subsection will be dedicated to paraphrase generation using T5 text-to-text transformed as proposed by Goutham [27]. The original pre-trained text-to-text model is fine-tuned to generate paraphrases by learning on Microsoft Research Paraphrase Corpus dataset [5], which contains 5800 pairs of sentences extracted from the Internet and their paraphrases. To run inference, we input an arbitrary encoded sentence and decode the results generated with the model. As an example we will try to generate five paraphrases of initial sequence: “While Steve Jobs was CEO, Apple developed some of its best products.” The generated paraphrases should hold the same meaning using different words. Additionally, we will add readability and polarity scores to see how they are altered by such paraphrasing. The scores of the initial sequence are:
As Apple’s chief executive, Steve Jobs led the way in developing some of its best products.

As Apple’s CEO, Steve Jobs developed some of its best products.

While Steve Jobs was chief executive, Apple developed some of its best products.

Jobs, while still CEO, developed some of Apple’s best products.

Steve Jobs, the new CEO of Apple, developed some of its best products during his time in office.

<table>
<thead>
<tr>
<th>Paraphrase</th>
<th>Readability score</th>
<th>Polarity score</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Apple’s chief executive, Steve Jobs led the way in developing some of its best products.</td>
<td>62.88</td>
<td>0.64</td>
</tr>
<tr>
<td>As Apple’s CEO, Steve Jobs developed some of its best products.</td>
<td>77.81</td>
<td>0.64</td>
</tr>
<tr>
<td>While Steve Jobs was chief executive, Apple developed some of its best products.</td>
<td>50.61</td>
<td>0.64</td>
</tr>
<tr>
<td>Jobs, while still CEO, developed some of Apple’s best products.</td>
<td>77.81</td>
<td>0.64</td>
</tr>
<tr>
<td>Steve Jobs, the new CEO of Apple, developed some of its best products during his time in office.</td>
<td>68.69</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 5.1: Paraphrase generation example using fine-tuned T5 text-to-text model

The generated paraphrases that we can see in Table 5.1 are indeed paraphrases of the original sentence. They all hold the same meaning by emphasizing that while Jobs was CEO of Apple, the tech giant developed some of its best products. In some sentences, “CEO” is replaced by “chief executive”, which once again share the same meaning, but the readability score is most likely lower due to higher syllables per word and word per sentence ratios. Additionally, the word “Steve Jobs” is replaced with “Jobs”, which once again generates a similar readability change. The last paraphrase in the example stands out a bit by including some additional text that slightly alters the meaning. By adding “the new CEO of Apple” we imply that the Steve Jobs was only recently named as the chief executive. This is a statement that cannot be interpreted from the initial sequence, which is why this paraphrase is probably the worst of the generated set. But despite the mentioned mistake the paraphrase holds much of the same meaning and can still be labeled as an acceptable replacement.
Regarding the output text evaluation metrics we can see that readability scores of paraphrases are not the same as the scores of the initial sequence. Considering the targeted value, we can either increase or decrease the output readability score by replacing sentences with their paraphrases. However, as it can be interpreted from the results, we cannot use this method to change polarity score, because the scores remain the same. The generated paraphrases mostly change the structure of the text, where altered words do not affect polarity score. The most important words such as “Apple”, “best products” and “developed” remain untouched. Nevertheless, we could expect that in some cases such word that will alter polarity score can be changed, but the crucial words, which affect polarity score the most, tend to remain the same.

**Testing influence of paraphrase generation on length and polarity**

We test our hypothesis by calculating an average change in polarity score of paraphrases of 400 random sentences from 400 random documents, one hundred from dataset of each considered publication type. The set of documents that we use for this hypothesis is a random subset of the documents used for the calculation of initial values of key text evaluation metrics in Chapter 3.

We generate five paraphrases for each sentence and calculate the mean relative difference. The mean relative difference of polarity scores between original sentences and their paraphrases is $0.91 \times 10^{-3}$, which means that we can use this method to alter solely readability scores and assume that polarity is not significantly affected. The same goes for length, which can also be affected by replacing sentences with their paraphrases. The mean relative difference of lengths between original sentences and their paraphrases is $0.11 \times 10^{-3}$. 
Chapter 6

Adaptation to publication types

As mentioned before, the goal of this thesis is to develop an algorithm that adapts texts to different target audiences. We described key text evaluation metrics that we will be adapting and the methods we can use to achieve the desired results. In this chapter the main contribution of our thesis is described. We will explain how the methods that were mentioned in the previous chapters will be used in the process of text adaptation. For better understanding we will generate flowchart diagrams that will show how the methodology works.

6.1 Structure

In this chapter, we will explain how the methods from previous chapters are put together for text adaptation to target publication types. We will show how our method works by first looking at the main flow of the methodology. Next, we will dig deeper into the main processes in our methodology. Firstly, we will look at the length adaptation process. Next, we will consider the flow of synonym replacement and finally, we will finish with the flow of automatic paraphrase generation.
6.2 Methodology

Being the main contribution of this thesis the flow of our algorithm to adapt texts to different publication types is explained in this chapter. Previously described methods are here represented in a single process that takes as an input the text that we want to adapt and the target publication type to which the text will be adapted. The input text is adapted iteratively, while trying to converge the key text evaluation metrics towards the initial values of targeted publication type.

In Figure 6.1 we can see the flow of our methodology. As mentioned before, it accepts input text and target publication type as inputs. Then we have to define the initial values of targeted publication type that we calculated in Chapter 3. We also define the number of iterations that we want our algorithm to execute. In the iteration we define the acceptable error $\epsilon$ due to variance scores in the datasets used to calculate baseline values of key text evaluation metrics. The iteration defines the change in text evaluation metrics that we want to apply. If the relative difference towards the initial value of some text evaluation metric is smaller than $\epsilon$, then we do not adapt this metric in the current iteration. In case relative_difference($t_i, m_i$) $< \epsilon$, we do not alter the text evaluation metric denoted by $i$, where $m$ is a dictionary of initial values of key text evaluation metrics of targeted publication type, $t$ is a dictionary of key text evaluation metrics of the text we are adapting.

Next, we start iterating through the main process by declaring index = 0, where index $\in \{0, 1, 2, \ldots, n_{\text{iterations}}\}$. In each iteration, we check if index $< n_{\text{iterations}}$ and if it is we adapt text. The text adaptation process consists of three different subprocesses. Firstly, we execute length adaptation subprocess, because it changes all of the key text evaluation metrics significantly. The subprocess is additionally explained in Subsection 6.2.1. After we finished length manipulation, we focus on text readability and polarity scores and adapt those with replacing words in text with optimal synonyms that minimize the relative difference between targeted and current values of text evaluation metrics. Further explanation of synonym replacement process
Figure 6.1: Flowchart of the main process of our methodology
is in continuation in Subsection 6.2.2. Lastly, we use a method that generally speaking alters only readability. Using paraphrase generation with text-to-text transformer, we slightly alter length and polarity as well, however as explained in Chapter 5, the change does not affect our results significantly. This process will be discussed in Subsection 6.2.3. All the described subprocesses replace the variable text with adapted texts that are output by these methods.

6.2.1 Adapt length

As mentioned in Chapter 4, length manipulation will play an important role in text adaptation, because of the main characteristics that define the chosen publication types. In this subsection, we will explain how the aforementioned methods are used in our methodology.

As we can see in flowchart in Figure 6.2, the method accepts the input text, initial values of targeted publication type and acceptable error $\epsilon$. Then we calculate the text evaluation metrics of text and their relative changes to initial values. If the absolute value of currently calculated relative change of length is smaller than $\epsilon$, then we can assume that the change is unnecessary and output the input text. Otherwise, we check if the current relative change is positive or negative. If it is positive, that means that the reference value, which is the initial length of targeted publication type calculated in Chapter 3, is smaller than the length of the input text, i.e., input text is shorter than the baseline value of targeted publication type. Thus we have to summarize our text. We do that using abstractive summarization technique, described in Chapter 4. Otherwise, we have to generate additional content using the natural language generation models we developed in Chapter 4. Before setting the output text, we have to beautify our text. Beautifying text simply means capitalization of the beginnings of sentences and entities and adding punctuation in the end, if necessary. This way we fix the mistakes that natural language generation models and summarization model produce during text adaptation.
Figure 6.2: Flowchart of the length adaptation process
6.2.2 Replace with synonyms

Length manipulation is followed by replacing words with optimal synonyms found in a dictionary. Using this method we will adapt both polarity and readability scores. In this subsection we explain the method in details.

From Figure 6.3 we can see that the methods once again accepts the input text, initial values of targeted publication type and acceptable error $\epsilon$. Additionally, we initialize the number of iterations, which denotes the maximum number of words that can be replaced by their synonyms. Paraphrasing paraphrases or finding synonyms of synonyms can quickly turn bad due to many meanings that words can hold. This is why we set the maximum number of replacements or iterations to $n = 15$, if we will be adapting text to social media, $n = 250$, if we will be adapting to research articles and $n = 100$ otherwise. Next, we initialize index $i$, where $i \in \{0, 1, 2, \ldots, n\_iter\}$ and a set of tokens from the input text sorted in the descending order by the sum of relative differences in polarity and readability scores, if we replace a word with a certain token. This way we minimize the relative difference of both scores. If $i$ is smaller than the number of iterations and the number of tokens, we search for the optimal synonym to replace the $i$-th token. We search through a dictionary to generate a set of synonyms of the word. Then we iterate through this set of synonyms and try to replace the word $w$ with the synonym $s$. However, since the synonyms found in the dictionary might not be in the same form as word $w$, we first have to fix that. In Python this can be done using several Python libraries, such as mlconjug (https://github.com/SekouD/mlconjug) for conjugating verbs and Natural Language ToolKit (https://www.nltk.org/) for form recognition using part-of-speech tags. If the optimal synonym is not set or if the sum of polarity and readability scores of a text, where we replace $w$ with $s$ are closer to the initial values of the target publication type, we can set the optimal synonym as $s$. After we find the optimal synonym, we replace the word $w$ in the input text with its synonym $s$. We then repeat calculating text evaluation metrics, increase $i$ and continue with the iteration until $i \neq n\_iter$. 
Figure 6.3: Flowchart of the synonym replacement process
6.2.3 Replace with paraphrases

After length and polarity score are fixed, only readability score remains. We will replace sentences with their paraphrases and thus try to adjust the remaining text evaluation metric. This subsection will explain how we implement that in this thesis.

Figure 6.4: Flowchart of the paraphrase replacement process
Just by looking at the chart in Figure 6.4, we can already see that it is almost the same as the one from the synonym replacement method, which is why we will not go into details of how it works as we did in the previous subsection. The list of sentences is not sorted, because we choose a sentence randomly from the list. Similarly as in Subsection 6.2.2, the number of iterations $n$ will be set to $n = 3$, if we adapt text to social media, $n = 50$, if we adapt to research articles and $n = 20$ otherwise. Randomly replacing $n$ sentences with their paraphrases will be enough to significantly improve the readability scores of our texts. Obviously, instead of synonyms we generate paraphrases and look for the optimal paraphrase. The optimal paraphrase replaces the sentence and minimizes the relative difference between text evaluation metrics of our text and initial values of target publication type.
Chapter 7

Results and discussion

As promised, the methodology is able to adapt texts to different publication types and thus we will generate results for each possible transition (e.g. from newspaper article to social media post). Additionally, we will present how the initial values of the key text evaluation metrics change throughout the execution and evaluate the efficiency and meaningfulness of the algorithm.

Now that we explained how our method works in theory, we will generate results to determine whether it can successfully adapt texts to different target audiences. In this chapter we will go through the chosen publication types and for each possible transition generate sample results. We will present graphically how the initial text evaluation metrics are transformed by the methodology and also show some examples of generated text. Due to time complexity of our algorithm, we will set the number of iterations of the outer loop to five. We will allow error of relative difference between initial values of target publication and text evaluation metrics of the input text: \( \epsilon = 0.1 \).

To get less biased results, we test our methodology by generating adapted texts of a subset of texts used for calculating initial values of different publication types in Chapter 3. This subset consists of 100 documents for each publication type (i.e., 400 altogether) that were randomly chosen from the main dataset. In the following sections, all the results are taken from the described dataset and adapted editions of input texts. All the charts in this
chapter show the average values of key text evaluation metrics of initial and adapted texts from this dataset grouped by publication types. Besides text evaluation metrics the adapted text also has to make sense. In this chapter we will also take a look at some examples of adapted text. Excluding social media posts, because some generated texts are very long, e.g. generated research articles usually contain about 4000 words, we will not interpret full texts but only introductory paragraphs of the generated texts.

7.1 Structure

In this chapter we will look at results and interpret them. We will start by interpreting results of all four publication types. In each of those sections, we will first evaluate how our method works by presenting results of text evaluation metrics before and after the methodology was applied. Then, we will go through the content of some generated examples and compare it with the initial text. We will do that for each possible transition of discussed target publication type. In such manner, we will first evaluate the results of social media posts. Then, we will continue with research articles. We will move on with official statement as the target publication type and conclude with news. Finally in this chapter, we will evaluate our method’s efficiency in the overall summary of the results we discussed in the previous sections.

7.2 Target publication type: Social media

In this section we evaluate results of adapting texts to social media from other publication types. We ran the method on 100 different cases for each publication type as described in the introduction of this chapter and calculated the mean of relative differences between the initial values of social media posts and both starting values of the input text as well as the final values after the adaptation. These means are shown in Figure 7.1.

From Figure 7.1, we can first see that in most cases we improve the relative
Figure 7.1: Absolute relative differences between text measures before and after adapting to social media

differences to the target values of text evaluation metrics. If we exclude transition to official statements, readability scores are close to the acceptable error $\epsilon$, which suggests that the whole method adapts readability very well. Even though results of polarity scores are not optimal, we still improved the initial situation of official statements by a lot, from which we can presume that the method can work in certain circumstances. Since polarity can only be altered with synonym replacement, which depends on the content and the number of available words, we can conclude that adapting polarity in
asocial media posts is harder to do due to lack of words with more optimal synonyms. The biggest improvements can be seen in length manipulation. Even though the length did not converge below $\epsilon$, the method significantly alters the length of the text using summarization technique. Moreover, there is no exact rule how many words should a social media post contain and because it usually consists of lesser words, the interval in which the result is acceptable is very small. On the other hand, social media posts with 50 or 100 words are generally acceptable and thus, we can say that we successfully adapted length to social media. In the following, we will look at the content of these results and see whether we could really post them on Twitter.

### 7.2.1 Social media post from official statement

Here, we will evaluate results of a few social media posts that we generated using our methodology, where the input texts were official statements.

*The coronavirus covid-19 pandemic is a global health and societal emergency that requires powerful immediate action by governments, individuals and businesses. Icc will regularly send updated.*

Figure 7.2: Example of text adaptation from official statement to social media post

The text shown in Figure 7.2 is appropriate to post on social media, if we only consider the key text evaluation metrics. However, we should also consider content, when it comes to text. For example, since no natural language generation was necessary in examples, because the initial text was much longer, we could not include Twitter characteristics in our text, such as emojis and hashtags. We have seen in Chapter 4 that our models include those characteristics in the generated text.

Moving on to the content the text is quite understandable. Obviously, a lot of data is lost in the process of summarization, but a very important message is generated as the adapted text. However, we can see that the last
sentence is cut off while summarizing. The original sentence says: “To aid this collective effort, ICC will regularly send updated advice to its network of over 45 million businesses so that businesses everywhere can take informed and effective action to protect their workers, customers and local communities and contribute to the production and distribution of essential supplies.”

We should add the word “advice” to the last sentence, which would make the sentence much more meaningful. Moreover, we would have to fix “Icc” to “ICC”. But otherwise the message of the text is clear and with these fixes we could imagine governmental institution posting this on Twitter.

### 7.2.2 Social media post from research article

Since research articles are much longer, it is difficult to find the corresponding sentences of the original article. This is why we will check just the topic of the research article and explain how the sentence might be related with the content.

*Among the wild bat paramyxoviruses identified in this study, 97% fell within a cloud of unclassified morbillivirus-related viruses.*

Figure 7.3: Example of text adaptation from research article to social media post

The original article from which the text from Figure 7.3 talks about highly diverse Morbillivirus-related Paramyxoviruses in wild fauna, which can also be seen from the output text. Some context should be additionally included in the social media post, but due to very low initial length of social media posts, the method correctly fails to do that.

As we calculated in Chapter 3, research articles are usually less readable. Regarding the structure, the original text holds much less readable data, higher words to sentences and syllables to words ratios, than the output text and regarding the key text evaluation metrics, the text was successfully adapted. If we included a link to the generated text that would justify the
usage of phrase “in this study”, we could post the text on Twitter.

### 7.2.3 Social media post from research article

The input text that was used to generate the text that is presented in Figure 7.4, talks about the first 1,000 cases in Canada. The sentence “Canada is officially in” indicates that Canada is a part of something, but the word that continues this sentence in the original text is “recession”. This significantly changes the meaning the text. However, we could also interpret this sentence by saying that it joined the countries with more than 1,000 cases. Additionally, it is hard to understand the first part of the second sentence, but we do conclude with a very informative sentence about the COVID-19 state in Canada that we could publish on Twitter.

Canada is officially in. Their jobs due to more than 1,000 state of matteresous state filled - roughly 6,000 tests and 11 confirmed cases of COVID-19 a report also details an estimated 79.

Figure 7.4: Example of text adaptation from newspaper article to social media post

### 7.3 Target publication type: Research articles

This section will be dedicated to showing results of text adaptation to research articles. In other words, we will show how efficiently our methodology adapts texts from other publication types to social media. Firstly, we will interpret the overall sample results using a similar chart to the one in previous section and then we will analyze sample transitions from other publication types to research article. Similarly as in the previous section, adapting to research articles was run on 100 different cases for each remaining publication type. The relative differences between the mean text evaluation metrics of
target publication type and text evaluation metrics of text before and after adaptation are shown in Figure 7.5.

![Figure 7.5: Absolute relative differences between text measures before and after adapting to research articles](image)

Figure 7.5: Absolute relative differences between text measures before and after adapting to research articles

Once again, we successfully improved most of the key target evaluation metrics. In this case we failed to adjust readability, most likely due to the same reasons as we failed to adjust polarity in the previous section. On the other hand, the synonym replacement method tries to lower the sum of relative differences of both polarity and readability, which we successfully did with each publication type. While some values do not fall under the chosen $\epsilon$ value, we can still say that we adapted the characteristics quite successfully,
because the relative differences after the text transformation are much lower overall. The method performed especially good with polarity scores, which is most likely a consequence of high variety of words that can be found in such long texts. When adapting articles of different publication types to research articles, we also found that length manipulation successfully falls below the $\epsilon$ value, with which we can conclude that our algorithm for natural language generation, which was used the most, in this case efficiently generates as many tokens as needed. In the following, we will see whether the content of the generated text is as appropriate as the number of tokens that are generated.

7.3.1 Research article from official statement

In the example in Figure 7.6, we can see the initial paragraph of input official statement and generated research article. We cannot say the text is completely meaningless, but it is obvious that words were replaced with their synonyms and some replacements were not adequate. For example, “round” was replaced with “moonlike”, which completely changes the meaning of the word. We can see that despite successful adaptation of text evaluation metrics, the method can wrongly alter the content, if such inappropriate synonyms and paraphrases exist that improve the text evaluation metrics and corrupt the content. In such case, we should also take into account the context of the words when choosing the synonym. However, as we mentioned in the introduction, computer cannot comprehend how words are connected as well as people do, which is why such errors sometimes occur.

In the continuation, the method generates a lot of additional text. In the beginning, the generated text is closely related to the context. It starts talking about a person from Ben-Gurion University of the Negev, which is a university in Israel, that was diagnosed with COVID-19 during an outbreak. The description of his state continues with:

“He is now receiving intensive care treatment for severe respiratory symptoms which include shortness of breath accompanied occasionally high fever
Azerbaijan is 1 of the 17 countries in the WHO European Region that is receiving personal protective equipment and laboratory testing kits from WHO – with a first round in February, and a second one currently being shipped. WHO deployed experts to the country for a mission to support preparedness and response. During the five-day mission the team visited national and regional hospitals, laboratories, discussed the national plan to detect and treat people with COVID-19. The team also worked with the national authorities to further develop the COVID-19 response roadmap. WHO will continue its support to Azerbaijan through its country and regional offices and by engaging international experts and partners in the country.

Azerbaijan is 1 of the 17 countries in the WHO European Region that is receiving personal protective equipment and laboratory testing kits from WHO – with a first round in February, and a second one currently being shipped. WHO deployed experts to the country for a mission to support preparedness and response. During the five-day mission, the team visited national and regional hospitals, laboratories, discussed the national plan to detect and treat people with COVID-19. The team also worked with the national authorities to further develop the COVID-19 response roadmap. WHO will continue its support to Azerbaijan through its country and regional offices and by engaging international experts and partners in the country.

Figure 7.6: Example of text adaptation from official statement to research article

and chills after returning into sleep following medical procedures such like MRIs [microsurgery].”

However, this does not look like a research article. It seems much more like a newspaper article. Since the text continues to talk about Coronavirus, we can assume that the model was successfully fine-tuned with such content and there is a possibility that an occurrence of similar text could be found in our training dataset (i.e., example a research article that discusses news related to COVID-19). We can look at another example of generated text, that does not correspond to COVID-19 or the situation in Azerbaijan.

“It has been well documented that there is no standard definition of real-time perception [3] so what does this mean? This paper aims towards providing some basic definitions via medical examples.”

This seems to be an introduction of an article that has something to do
with medicine, but might not be the most appropriate to include it in our article. On the other hand, in this sentence, we can find one of total 22 citations that were generated in the article. Our method will most likely perform worse, when generating lots of additional data due to the inability to take into account the whole text. Furthermore, the model could be improved by additional fine-tuning, for which we would require a computer with higher computability capabilities. Even though, the generated content does not completely relate to the input text, we have to take into consideration that research articles are usually written by highly educated people and even they sometimes struggle with the content. Our method generated more than 4000 additional words, which include citations that are one of the main characteristics of research articles. The words are reasonably put into sentences as can be seen from the shown examples.

7.3.2 Research article from social media post

In Figure 7.7, we can see how a tweet is transformed to a research article. While we successfully generate more than 4000 words, the text does not look like research article. As the content of input text is not academic, we can see that our method fails to transform text into a text that would look like a research article. On the other hand, no reasonable person would ever try to write an article from such social media post. Our method successfully continues the text by adding the content colored with green in Figure 7.7. The post is an appreciation of those close to the publisher and our method interestingly relates this to mental health. As we can see in texts colored with red, mental and psychical health are then considered throughout the whole text, which cannot go overlooked.

7.3.3 Research article from newspaper article

Transforming newspaper article to research article, firstly, requires generation of additional text. We will evaluate the research article that was generated
Figure 7.7: Example of text adaptation from social media post to research article

from a newspaper article that talks about the situation of Coronavirus in Brazil and the impact of their dictator Jair Bolsonaro. The generated text fails to continue the topic, because our model cannot take into consideration longer texts, but merely last few sentences. Consequently, the method continues the input text with content related to Fidel Castro, who was a Cuban dictator. In continuation we can also find the following text:

“In January 2016, UNAIDS reported two cases where doctors were injured while treating suspected dissidents at hospitals run by potential unitary with Cuba International Medical Center — a January 2016 case where ...
ically injured democratic man died after surgery on his left leg; and another patient who had been transferred there hours earlier due an emergency condition similar not only directly caused by prolonged exposure. Two more bodies remain unidentified following their embalming procedures this week.”

We can see that not much content is related with Coronavirus and we conclude that our methodology does not perform optimally, in case the targeted length is much bigger than its initial value. However, the method seem to generate understandable sentences that can be related to the last sentences of the prompt sequence. In our example, the generated text is first related to Fidel Castro, who is a dictator like Jair Bolsonaro. Then we relate Fidel Castro to Cuba and its incident of HIV patients in January 2016. Even though the generated text does not hold a common thread throughout the whole content, the ability to form a chain of connections between paragraphs in generated text is very promising.

7.4 Target publication type: Official statements

In this section we will evaluate results of our methodology by adapting texts to official statements from other publication type. We can differ official statements from other publication types especially by high polarity score and low readability. First, we will see the mean values of relative differences between the initial values of official statements that were calculated in Chapter 3 and the key text evaluation metrics of the input texts before and after text adaptation. The relative differences hold the mean values of 100 generated examples from each publication type.

Based on Figure 7.11, adapting texts to official statements proved to be the most efficient as we successfully adapted all of the key text evaluation metrics by lowering the difference to the initial values of official statements that were calculated in Chapter 3. We also managed to get all the values quite near the acceptable value \( \epsilon \), which means that all the text metrics
Figure 7.8: Absolute relative differences between text measures before and after adapting to official statements

were converging towards the desired values. Why was adaptation of key text evaluation metrics to official statements so efficient? A possible explanation could be that official statements may be more lexically diverse, which is why more synonyms may be found due to multiple meanings that a word can hold. In the following, we will look at some examples and evaluate whether the generated content is meaningful or not. Because the generated texts are very long, only a few sample results from each generated text will be shown.
7.4.1 Official statement from social media post

In Figure 7.9 we can see an example of transforming a tweet to an official statement. It uses paraphrase generation to generate “Authentics: Politians and bussinesses want a story like this” from “Politicians and bussinesses want your money”. While the paraphrase is not the most accurate, it very well connects with the rest of the text. Additionally, we replace “bad” to “unspeakable”. This signifies that the original text either has higher readability or polarity score than the initial values of social media posts, because the word is usually used as more negatively-oriented, which means that it will decrease polarity and it contains more syllables, which means that it will decrease readability. The shown first sentences in generated official statement are nicely connected to the input text. Additionally, the generated text is much more appropriate for a governmental institution to publish than the input sequence. If we changed the input text more efficiently, we easily publish the text in a press release.

#coronavirus #covid19 deaths continue to rise. It's almost as bad as it ever was. Politicians and businesses want your money.

#coronavirus #covid19 deaths continue to rise. It's almost as unspeakable as it ever was. Authentics: Politicians and businesses want a story like this. Co/2Jh6cm7yj0 — it (@it_com) February 2, 2017 Dr. Sanjay Gupta of CNN medical center in New Delhi says the virus has spread through air across Assam: “more than 1,200 people have been hospitalized with respiratory light-headedness or severe skin rashes after being exposed”, (source.) His hospital reports that there are about 937 cases confirmed by them so far for every 100,000 patients treated at his facility... The WHO is currently monitoring two additional countries — nepal (with 1076 radical case count from march 15-17) and bangladesh (1116). We believe these numbers may need further updates when they become available,” said Dr. Sushil Kumar Panicker, director general & chief medical officer of world health organization ...

Figure 7.9: Example of text adaptation from social media post to official statement
7.4.2 Official statement from newspaper article

Since newspaper articles are often adapted from official statements, the changes between the input newspaper article and the generated official statement are not as obvious. If we consider the initial values calculated in Chapter 3, we can also see that the values differ only in readability and polarity scores, which are hard to find in longer texts. As an example we will see how our method adapts first few sentences of the input text.

The latest: The lives behind the numbers: What we know about the first 1,000 COVID-19 deaths in Canada. Canadians have lost more than $1.2 million to COVID-19 scams. Canada is officially in a recession, C.D. Howe Institute says. Nova Scotia is easing some of its COVID-19 restrictions.

Canadians have missed more than $1.2 million dollars a twelve months spent on cheats involving television commercials, Canada is officially in a economic condition, C.D. Howe Institute add : Nova Scotia is easing some of its covid-19

Figure 7.10: Example of text adaptation from newspaper article to official statement

From Figure 7.10, we can see how paraphrases are formed from the initial newspaper article. The word “lost” is replaced with “missed” and “on cheats involving television commercials” replaces “COVID-19 scams”. Both alterations are not optimal, however, the latter additionally corrupts content, which is why our method should not have included it. In this example we can see how automatic paraphrasing can also choose inappropriate paraphrases and adds additional content such as in our example “a twelve months spent”, which is also inappropriate. While paraphrase generation can efficiently alter readability, it is difficult to include content in paraphrase generation and can thus replace the sentences inadequately. Once again the reason is the inability of our algorithm to understand the content as people do and thus decide what is appropriate and what not. On the other hand, we saw that in some cases the generated paraphrases are appropriate and consequently efficient in readability alteration.
7.4.3 Official statement from research article

Integration of functional electronic devices on rigid surfaces can enable smart structures with sophisticated functionalities and applications [1] [2] [3] [4] such as bio-integrated electronics in healthcare systems, dynamic structural health monitoring in aerospace applications and miniature radar devices for defense applications [1] [2] [3] [4]. A critical enabler of these applications is the fabrication of electrically conductive 3D interconnects conformally on to 3D surfaces [2, 4, 5]. These interconnects can be combined with pick-and-place methods for integrating other active devices. Past work has shown several methods that can fabricate such conformal interconnects. In transfer printing, the ability to conform to complex surface features is limited to smooth curved surfaces due to the inherent need for a mechanical handling system to transfer the interconnects embedded on planar elastomer to the target 3D surface [3, 6].

Integration of functional electronic devices into rigid structures can enable sharp defenses with well-informed functionalities and applications. Bio-integrated electronics in healthcare systems, self-propelled structural health monitoring in aerospace applications and miniature radar devices for defense applications [1] [2]. Such devices can be made electrically conductive by bonding 3D interconnects to 3D surfaces, forming a layer of thin film... A key feature of this approach is landscaped conformance to sharp surface features – wider material window due to deficiency of heating of the 3D object... These interconnects can be combined with pick-and-place methods for integrating other active devices. Transfer printing leaves the magical form to conform to complex surface features to smooth curved surfaces due to the inherent need for mechanical handling system to transfer the interlinks embedded on planar elastomer to target 3D surface.

Figure 7.11: Example of text adaptation from research article to official statement

Here, we will evaluate the sections that can be extracted from the origin official statement and the adapted research article. In Figure 7.11, we emphasized interesting paraphrases with yellow. We can see that “interconnects” is replaced with “interlinks”, most likely to increase readability. Even more interesting is the replacing “A critical enabler of these application is the fabrication of electrically” with a paraphrase “Such devices can be made electrically”. While the paraphrase loses some meaning, the main concept that is described in the continuation remains the same. With green, we emphasized the injections of related text using abstractive summarization. This was the primary reason, why we chose this method over extractive approach.
It constructed two sentences from one long sentence, which can significantly improve readability as well.

Even though the output version of the section loses some meaning, we can agree that the initial text is much less similar to an official statement in comparison to the adapted text. In approximately the same amount of words, it expresses more information, which is essential when generating shorter texts such as we did in this example.

### 7.5 Target publication type: News

In this section we will evaluate the results of the remaining publication type, news. News are generally as long as official statements, but they have lower polarity score and are also more readable. Otherwise, as we will see in this section, newspaper articles can be very similar to official statements and are often just a discussion about governmental press releases. In the continuation, we will look at relative differences between the initial values of newspaper articles and both starting values of the input text as well as the final values after the adaptation. Similarly as in the previous sections the relative difference hold the mean values of 100 generated examples for each publication type.

We can see in Figure 7.12 that we failed to improve all the key text evaluation metrics. When adapting the readability scores, the values after adaptation were worse than they were before the method was executed. This is a consequence of a very small initial relative difference between readabilities. Since the initial difference of polarity scores is higher and we try to find the optimal synonym that improves the sum of absolute values of readability and polarity, the synonyms that improve polarity affects the relative difference more and will consequently be prioritized. This is why polarity is improved in the output text much more than readability. Additionally, the initial relative difference of polarities is slightly higher and the sum of final differences is still more than halved after adaptation, which is why we can
conclude that our synonym replacement method worked efficiently. Furthermore, we achieved optimal results in length and polarity adaptation. In the following, we will see how meaningful is the generated text.

### 7.5.1 Newspaper article from official statement

In the following, we will see how all three key text evaluation metrics are manipulated when adapting official statements to newspaper articles.

In Figure 7.13, we can see that the paraphrase excludes the first part of the sentence. This slightly changes the message of the text, however, if
to combat the coronavirus COVID-19 pandemic, the International Chamber of Commerce (ICC) and the World Health Organization (WHO) have agreed to work closely to ensure the latest and most reliable information and tailored guidance reaches the global business community.

The International Chamber of Commerce and the World Health Organization (WHO) have agreed to work closely to ensure the latest and most reliable information and tailored guidance reaches the global business community.

Figure 7.13: Example of readability adaptation from official statement to newspaper article

we published a text in this form, we would perfectly understand what is the activity in which organizations will work together. Even though that in some cases this will work against us, performing this task makes paraphrase generation capable of a human-like restructuring, i.e., we might do this ourselves as well. In this case we also manipulate length.

The COVID-19 pandemic is a global health and societal emergency that requires effective immediate action by governments, individuals and businesses.

The COVID-19 pandemic is a global health and societal exigency that requires successful immediate action by governments, individuals and businesses.

Figure 7.14: Example of polarity adaptation from official statement to newspaper article

From example in Figure 7.14 we can see that we replace the word “emergency” with its synonym “exigency”. In this case this is an appropriate switch that increases polarity score. Same goes for replacement of “effective” with “successful”. The meaning of the text remains the same and polarity score is increased by a word that has more positive meaning.

Lastly, in Figure 7.15 we can see another example of efficient paraphrase generation that increases readability by shortening the sentence. The method similarly as in Figure 7.13 efficiently removes a subsentence.
To aid this collective effort, ICC will regularly send updated advice to its network of over 45 million businesses so that businesses everywhere can take informed and effective action to protect their workers, customers and local communities and contribute to the production and distribution of essential supplies. ICC will also contribute to enhancing information flows on the coronavirus outbreak by surveying its global private sector network to map the global business response.

To aid in this collective effort, ICC will regularly send updated advice to its network of over 45 million businesses, so that businesses everywhere can take informed and successful action to protect their workers, customers and local communities.

Figure 7.15: Example of text adaptation from official statement to newspaper article

7.5.2 Newspaper article from social media post

The generated text from the example in Figure 7.16 generates a paraphrase that is very imaginative, but the start of the paraphrase “Under the new law, you no longer have to wear” means exactly the opposite as the input text message suggests. If we exclude that fact, the generated text looks a lot like a newspaper article. Text generated with NLG model is related to the content and well formed. Overall, the introduction that is shown in Figure 7.16 looks like an exemplary introduction to a newspaper article.

You now have to wear face coverings when out shopping - this includes a visit to your local Community Pharmacy...

Figure 7.16: Example of text adaptation from social media post to newspaper article
7.5.3 Newspaper article from research article

For comparison we will use the same text to generate newspaper article as we did in Figure 7.11.

Integret of functional electronic devices into rigid structures can enable sharp defenses with well-informed functionalities and applications. Bio-integrated electronics in healthcare systems, self-propelled structural health monitoring in aerospace applications and miniature radar devices for defense applications [1][2]. Such devices can be made electrically conductive by bonding 3D interconnects to 3D surfaces, forming a layer of thin film. A key feature of this approach is landscaped conformance to sharp surface features - wider material window due to deficiency of heating of the 3D object. These interconnects can be combined with pick-and-place methods for integrating other active devices. Transfer printing leaves the magical form to conform to complex surface features to smooth curved surfaces due to the inherent need for mechanical handling system to transfer the interlinks embedded on planar elastomer to target 3D surface.

Figure 7.17: Example of text adaptation from research article to newspaper article

Curiously, the method produces the same result as shown in Figure 7.11. Length is changed similarly, because the target values are almost the same. But why are the readability and polarity not changed? It is probably changed in other sections of the text. Another reason might be that the final relative difference of readability scores is much higher in generated newspaper articles, than it is in generated official statements. This simply means that no better solution for this section was found in the given number of iterations.

7.6 Overall summary

Now that we have seen our method at work, we can safely say that we obtained satisfactory results in adaptation of the considered key text evaluation metrics. On the other hand, the results are not as perfect when it comes to content. Our method has its drawbacks such as generating lots of additional content, which often results in an unconnected text. Additionally, synonym replacement and paraphrase generation can incorrectly replace original sentence or word, where the paraphrase or synonym changes the meaning but
proves to be efficient when adapting text evaluation metrics, if there exist such synonyms that are more appropriate to use for a particular target audience. Nevertheless, our methodology generated a few sequences that can be published for target audiences right away without any change and lots of texts would only require minor corrections.

While our method successfully adapts key text evaluation metrics and generates some well-adapted texts, it can be seen from some texts, that the results do not necessarily hold main characteristics of target publication type that cannot be discretized, such as hashtags, emojis, citations and most importantly appropriate content. We saw that not every social media post can be adapted to research article and while with the right topic, we might achieve better results, we have shown what happens if the topic is not appropriate. To clarify, if we wanted to generate a research article of a tweet about appreciation of friends, we would end up generating lots of senseless data, because no such research articles exist and consequently our model cannot comprehend correctly what we want as the output, because we did not input the appropriate training data.

Our methodology obviously performs better for some audience transformations. Our methodology performs very well if the text does not contain words of which optimal synonyms and optimal paraphrases corrupt the message. The transitions that gave us the best results were:

- when adapting any publication type to social media posts,
- when adapting research articles to newspaper articles,
- when adapting official statements to newspaper articles,
- when adapting research articles to official statements,
- when newspaper articles to official statements.

To conclude this chapter, we are satisfied by the benchmarking results that our method produced in adapting key text evaluation metrics. The methodology produces some fascinating content and can thus be used as a baseline for further text adaptation to target audiences.
Chapter 8

Conclusions

In this thesis we developed a methodology that adapts texts to context. The methodology focuses on three text evaluation metrics: length, readability and polarity of the text. We calculated the initial values of four different publication types that represent different target audiences. After that our method iteratively adapts text to these initial values based on the targeted publication type by adapting texts to the chosen key text evaluation metrics. We alter length using abstractive summarization with T5 text-to-text transformer and fine-tuned natural language models for generating additional text. Polarity and readability scores are additionally adjusted by replacing words with more appropriate synonyms. In the end of each iteration the method replaces sentences with appropriate paraphrases generated with fine-tuned T5 text-to-text transformer.

To conclude, the problem of text adaptation proved to be much more difficult than expected. While we found text evaluation metrics that define different publication types, in some cases adjusting these measures is not enough. If we want to generate longer sequences (more than 2000 tokens) with our NLG models, we find that the generated text is not connected and while we can find a chain of related topics of subsections, in some cases it is hard to define the common thread that is held throughout the whole generated text. Additionally, when such synonyms and paraphrases exist that
corrupt the content but improve the relative differences to the targeted values of key text evaluation metrics, the methodology will replace existing words and sentences with corrupted content. However, despite these drawbacks, we generated lots of results that reflect the targeted publication types and even more results that would require only minor changes to be completely acceptable.

Additionally, we successfully developed a methodology that adapts the key text evaluation metrics to targeted values. The obtained results can be used as a baseline for further research. The methodology improves the text evaluation metrics of the input text in nearly all of the shown examples. Nevertheless, as we described in the previous paragraph, sometimes this is not enough. We conclude this thesis with satisfactory results of both content of generated texts and their values of key text evaluation metrics.

Our ideas for further work include improvement of natural language generation model, where the pre-trained model that we used should be trained on longer texts, so that we could generate text based on longer prompts and thus make sure that we hold the common thread throughout the whole text. Additionally, we should also work towards developing a method, with which we could decide whether synonyms or paraphrases corrupt the message of the text. Word embedding can be used to represent the context of the text and we could use it to determine whether the synonym fits the current context or not. Another way to adapt text to context would be to create a dataset of texts, where each row would hold different versions of the same text and each version represents the text written for different target audience. This way we would be able to teach text-to-text models to adapt text to context and it could also consider patterns that are not obvious to the human’s eye.
Bibliography


