Carnegie Mellon University Qatar

Narrative Summarization in the Domain of Finance

Dietrich College Senior Honors
Information Systems, CMU Qatar

Prepared by: Samir Abdaljalil

Advisor: Prof. Houda Bouamor

Abstract

The number of electronic text documents is growing and so is the need for automatic text summarizers. In the finance domain, documents can be quite long, averaging at approximately 180 pages. This creates a need for finding efficient ways to make use of technology to leverage the existence of these textual datasets. This goes hand in hand with the pressing need to make investment/financial decisions in a fast manner to ensure maximized financial gain. However, exhaustive reading of financial documents is extremely laborious. Hence, Automatic summarization methods could greatly simplify this task. In this work, we present several approaches for summarizing the qualitative sections of annual reports using extractive summarization, Natural Language Processing (NLP), and Machine Learning techniques. We investigated multiple approaches under two different types of explorations, sentence-based summarization and a section-based summarization tailored to the structure of financial reports. We then evaluated the quality of the summaries using an existing dataset of annual reports published by FNS-2020 shared-task that consists of annual reports by British Firms belonging to the London Stock Exchange.

Acknowledgements

This thesis had a lot of aspects to it that could not have been achieved without the help of the supportive people within CMU and beyond.

First and foremost, the valuable guidance and support that I received from my advisor, Dr. Houda Bouamor, was one of the main reasons why I was able to successfully complete this rigorous, yet rewarding, research. She allowed me to experience what a true research project is like, with its many unexpected twists and turns, but was also there to guide me through the whole process whenever I needed her support. Coming into this research, Natural Language Processing was merely an interest of mine, and her involvement in this project made it possible for me to come out of this experience with a lot more knowledge than I came in with.

Furthermore, I would like to thank Dr. Mahmoud El-Hajj and Ahmed AbuRa'ed who supported me especially with obtaining the data and evaluating my systems appropriately. Their support and cooperation was integral to the proper completion of this research, and I am highly appreciative of the time they took to meet with me and/or provide me with feedback.

I would also like to thank Dr. Serkan Akguc for his support in navigating the Finance aspects at the beginning of this research. Meeting with him helped me gain even better understanding of the importance of my research to the Finance field, and its potential applications in the industry.

Finally, the constant support and encouragement that I received from my family, especially my parents, and friends, not only throughout this thesis, but throughout my whole degree, was one of the reasons that I was able to achieve the results I have achieved.

Table of Contents

Abstract 1

Acknowledgements 2

1. Research Questions and Significance 4

2. Related Works 7
- 2.1. The role of annual reports in practice 7
- 2.2. Extractive Summarization Approaches 7
- 2.3. Summarizing Financial Documents 9

3. Data 10
- 3.1. Data Preprocessing 11

4. Methodology 13
- 4.1. BERT for Text Encoding 13
- 4.2. Sentence-based Summarization Approaches 15
  - 4.2.1. Sentence-based summarization as a Classification task 15
  - 4.2.2. CENTROID-SENTENCE-BASED method 17
- 4.3 Section-based Summarization Approaches 18
  - 4.3.2. SECTION-COSINE 18
  - 4.3.3. SECTION-CLUSTERING 19
  - 4.3.4. WEIGHTED-SECTION-CLUSTERING 19

5. Evaluation and Results 21
- 5.1. Intrinsic evaluation 21
- 5.2 Extrinsic Evaluation 23

6. Conclusion and Future work 25

1. Research Questions and Significance

As technological resources are evolving, different domains are starting to adopt technology, and make use of it in a way that can make certain aspects of that domain more efficient. The digital world has become more complex and crowded with massive volumes of digital data. In 2018, the size of the indexed World Wide Web is over 5.22 billion pages (Kunder, 2018), spread over 1.8 billion websites (Fowler, 2018). As the number of electronic text documents is growing so is the need for an automatic text summarizer. This ideology applies to the domain of finance. In this domain, reading and understanding long texts is time and effort-consuming. Consequently, automatic text summarization can be seen as a viable solution. It can decrease the time taken to summarize huge texts and create summarized reports.

In the finance domain initially, there seemed to be a lack of technological resources that enabled individuals in the field to perform their jobs efficiently. However, over the years the emergence of the Fintech industry has disrupted the way individuals operate in the Finance domain. In fact, according to PricewaterhouseCoopers, "Funding of FinTech startups has increased at a compound annual growth rate (CAGR) of 41% over the last four years, with over US$40 billion in cumulative investment." (PWC, 2018) According to JP Morgan, the Banking and Securities industry has been widely investing in Artificial Intelligence (AI) applications, a good example being the usage of news sentiment analysis for automatic investment (2020). Furthemore, AI applications have been used to aid with managing quantitative data in the financial domain to provide automated credit-scoring for prospective bank clients by analyzing social media activity and their search history (Leidner, 2019).

This sudden surge of interest and investment in the FinTech industry, and technology in general is understandable, as from this domain comes a large amount of data created by different firms. Such data comes in many different forms. For instance, firms and businesses worldwide use a number of different methods to communicate with their shareholders and investors and to report to the financial markets. Such methods include annual financial reports, quarterly reports, preliminary earnings announcements, conference calls and press releases (El-Haj, Rayson, & Moore, 2018).

Most of these documents are published in a PDF file format including figures, tables, numbers, and most importantly text(narratives). Financial documents can be quite long, averaging at approximately 180 pages (2020). An example can be seen below[1]. This creates a need for finding efficient ways to make use of technology to leverage the existence of these textual datasets, especially to extract the most relevant information from different key sections. For instance, Investment banking firms are investing in experienced financial analysts to be able to find and analyze market reports and other related news to aid them in excelling against their competitors. In many instances, however, those analysts are bombarded with hundreds, if not thousands, of pages to read on a regular basis. This means that they are most likely going to not read everything presented to them, resulting in them overlooking crucial information that might benefit the company. This is where summarization systems can be useful, as these analysts would be able to go through those summaries of reports and other related documents, and derive appropriate market conclusions (Khant & Singh Mehta, 2018).

Automating these financial document analysis processes would go hand in hand with the idea of Time Value of Money (TVM) to investors, since the value of money fluctuates constantly, and the need to make investment/financial decisions in a fast manner is essential to ensuring maximized financial gain (Chen, 2009). With the rise of the application of AI in automating certain processes, the use of NLP peaks the interest of professionals in the Finance domain, since these financial reports largely consist of textual data. A substantial amount of research has been conducted on the use of NLP and text-mining techniques in the financial domain by looking into sentiment analysis and information extraction techniques applied to Financial news, and looking into applying extractive summarization techniques to Financial news as well (Filippova, Surdeanu, Ciaramita, & Zaragoza, 2009). However, research on how

[1] https://www-file.huawei.com/-/media/corporate/pdf/annual-report/annual_report_2019_en.pdf?la=en

to apply NLP and Machine Learning techniques to analyze and summarize other textual data-sets in the finance domain like Annual reports has not seen major development yet, due to the extremely unstructured nature of those reports.

There are two general approaches to automatic summarization: extraction and abstraction. Extractive summarization methods involve the extraction of important key phrases and sentences from a source document to create the new summary. It consists of first identifying the important sentences or phrases from the original text and extract only those from the text. Then after applying a ranking function, only a subset of those sentences would generate the final summary (V. Gupta & Lehal, 2010). Therefore, identifying the right sentences for summarization is of utmost importance in an extractive method. On the other hand, abstractive summarization involves producing "a generalized summary, which conveys information in a precise way that generally requires advanced language generation and compression techniques." (Moratanch & Chitrakala, 2016) In other words, it performs summarization by taking the most important text from the original document, and puts it through an abstractive summarization system that rewords that text, and creates new sentences that are to be included in the final summary.

In this research, we will be exploring extractive summarization approaches due to the lack in terms of datasets: there is no dataset of annual reports and their equivalent abstractive summaries publicly available. Building such a dataset is expensive in terms of time and funding.

This brings us to the following research questions:

1- How can we use NLP and Machine Learning techniques to build automatic financial narrative extractive summarizers?

2- Which data could be used to build a model for automatic summarization of Annual Reports?

3- How do we evaluate the usefulness of the summaries produced and techniques explored?

2. Related Works

In this section, we present an overview of the research work dealing with financial documents. We also present the existing systems for summarizing annual reports.

2.1. The role of annual reports in practice

Within the finance domain, annual reports have been heavily studied in several research journals. Their purpose and role in a company are constantly being explored in the Finance research field. For instance, Stanton & Stanton (2002), discuss the importance of a financial annual report to a firm's activities. They define what role that plays, by essentially stating that a financial annual report is a detailed account of a firm's activities for the last financial year. Their main aim is to be shared with relevant shareholders and investors to keep them up to date on the firm's financial endeavors during the previous year. Furthermore, they reference studies that analyze the different sections of an annual report, and how the content is tailored to each section depending on the type of the firm, or the audience that it is intended for (Stanton & Stanton, 2002). In addition, Ghazali & Annum (2010) look at the true usefulness of having a corporate annual report for companies, by looking within the context of Malaysia. They go on to state the importance of an annual report and its effect on a company's image, as in some cases as well, an annual report is used as a chance for the company to gain some traction and is used as an effective marketing tool. Most annual reports incorporate an overview of the types of services or products offered by the firm, and which markets the company serves. Furthermore, it allows prospective and/or current customers to build a sense of confidence towards the brand as they are able to clearly view any changes in operation, and the profitability profile of the company as well (Ghazali & Annum, 2010).

2.2. Extractive Summarization Approaches

Automatic narrative extractive summarization has been studied in several research works. Many systems based on various approaches have been introduced in the literature. Leidner (2019) explores the different Artificial Intelligence/NLP techniques that are used in text summarization. He writes about the specific application of NLP in summarizing documents in the Finance and regulatory domain. He discusses the typical quality dimensions of a summary, for instance: whether it is language-agnostic or language-dependent, or whether cross-domain robust or domain-specific etc. He then moves on to discuss the different methods used in summarizing financial documents by looking into heuristic, statistical, and deep learning applications such as neural methods (Leidner, 2019). Abujar et al. (2017) propose a heuristic approach to extractive summarization of Bengali text. By identifying tokenized word frequency scores, and deducing sentence scores, which would identify the most important sentences to be included in the summary. Another approach to this task is applying graph-based methods to automatic summarization. (Abujar et al., 2017) Xu et al. (2013) propose a graph-based extractive summarization method for multi-tweet summarization, by making use of named entities, and frequency of topics discussed within the tweets (Xu et al., 2013). Similarly, Mihalcea (2004) applies the graph-based TextRank algorithm to single-document summarization of news articles, which ranks sentences based on their connections and similarity scores between other sentences within the news article (Mihalcea, 2004). Furthermore, Query-based automatic summarization has also been explored. In fact, the most closely related work to the proposed research, in regards to Financial documents, would be research conducted by Fillippova et al. (2009), which presents an extractive summarization system for summarizing financial news. They fill this gap by introducing a simple algorithm that takes a company name as input, and retrieves any financial news regarding that company posted on Yahoo News, and ultimately ranking sentences in terms of importance or relevance (Fillippova et al., 2009). Furthermore, Berger & Mittal (2000) employ a statistical approach to query-based extractive summarization, by making use of frequently-asked questions documents found on websites, as "[They] view each answer in a FAQ as a summary of the document relative to the question which preceded it." (Berger & Mittal, 2000) Whereas, other researchers have applied Artificial Intelligence techniques to explore the extractive summarization task. Chuang & Yang et al. (2000) applied this on U.S Patent and Trademark documents by segmenting sentences using clauses, and identifying a total of 23 features of those segments, such as average term frequency, and paragraph number. Once those features were determined, they explored several machine learning algorithms, including DistAI (Chuang & Yang et al., 2000). Introduced by Yang et al. (1999), DistAI is described as "Multi-layer networks of threshold logic units (TLU) [that] offer an attractive framework for the design of pattern classification systems." (Yang et al., 1999) However, with all this literature addressing the extractive summarization task, its applications in the Finance industry are very limited, and finding its relevance to lengthy financial documents such as annual reports is a very beneficial task for both the Finance and the NLP industries.

2.3. Summarizing Financial Documents

The application of summarisation and natural language processing techniques in general has promising applications in the financial domain (El-Haj et al., 2019). Recently, statistical features with heuristic approaches have been used to summarise financial disclosure texts (Cardinaels et al., 2019), generating summaries with reduced positive bias and leading to more conservative valuation judgements by investors that receive them. Furthermore, the financial narrative summarisation task (El-Haj, 2019) of the Multiling 2019 workshop (Giannakopoulos, 2019) involved the generation of structured summaries from financial narrative disclosures. It aimed to provide researchers in the field of NLP with a platform to explore the different approaches of extractive automatic summarization to UK annual reports, while also "[Demonstrating] the value and challenges of applying automatic text summarization to financial text written in English, usually referred to as financial narrative disclosures." (El Hajj, 2019). In fact, this task was extended to create the FNS 2020 Shared Task included in the FNP-FNS workshop held in 2020. Several systems have been introduced. Zheng, Lu, & Cardie (2020) proposed their 'SUMSUM' system that involved splitting the annual reports into their relative sections by parsing the Table of Contents and then applying a BERT-based classifier to determine which section to include in the final summary. It is important to note, however, that any BERT model can process a maximum of 512 tokens at once (Zheng, Lu, & Cardie, 2019). In this case, despite the length of each section, Zheng et al. (2020) only appropriately processed the first 512 tokens of every section, meaning that their BERT representations aren't a 100% accurate representation of the section as a whole. On the other hand, Azzi & Kang (2020) implemented a similar approach with extracting the Table of Contents (TOC) from each annual report, they made use of a Convolutional Neural Network (CNN) binary classifier using Keras, that classified all titles as either narrative or not, based on their presence in the reference summaries provided (Azzi & Kang, 2020). On the other hand, Singh (2020) uses a different approach, that is based on combining both extractive and abstractive summarization methods by exploring "Pointer networks to extract important narrative sentences from the report, and then T-5 is used to paraphrase extracted sentences into a concise yet informative sentence." (Singh, 2020) .

3. Data

In this work, we focus on annual reports produced by UK firms listed on The London Stock Exchange (LSE). The dataset is built for the FNS Shared task and was made publicly available. The task dataset has been extracted from UK annual reports published in PDF file format. UK annual reports are lengthy documents with around 80 pages on average, some annual reports could span over more than 250 pages, while the summary length should not exceed 1000 words. The training set includes 3,000 annual reports, with 3-4 human-generated summaries as gold standard. For the evaluation process the test set of 500 files were provided.

As the reports are provided in PDF file format, extracting structure is a challenging task. El-Haj et al., 2018b, 2019b used the UK annual report's table of contents to retrieve the textual content (narratives) for each section listed in the table of contents. Section headings presented in the table of contents are used to partition retrieved content into the audited financial statements component of the report and the "front-end" narratives component, with the latter sub-classified further into a set of generic report elements including the Chairman's Statement, CEO Review, the Governance Statement, the Remuneration Report, and report's Highlights. For the creation of this dataset a number of 3,863 annual reports was used. Table 1 shows the dataset details:

Type	**Training**	**Validation**	**Testing**
Full Annual Reports	3000	363	500
Gold Summaries	9730	1250	-

Table 1: FNS-2020 Dataset Distribution

3.1. Data Preprocessing

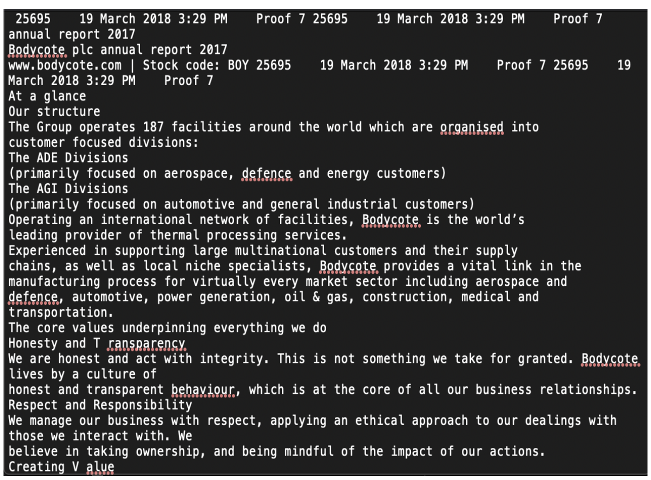
Data preprocessing includes cleaning the data from any unwanted (noisy) information that is not necessarily needed when it comes to generating the narrative summaries. Due to the previous nature of the texts being in PDF format, the organization and structure of the resulting TXT files is extremely messy. “Most of the established PDF to text conversion products on the market (i.e., pdf2text) generate highly noisy unstructured texts containing abbreviations, non-standard words, false starts, missing punctuation, missing letter case information, and other text disfluencies.” (Azzi, Bouamor, & Ferradans, 2019) Most sentences span across multiple lines, without a clear indication of where a specific sentence starts or ends, with lots of empty lines throughout the document. Since extractive summarization is solely based on the information within the texts, it is important to be able to identify and process individual sentences or sections. This process was extremely crucial in ensuring that the data was viable for automatic summarization, which is why it took a large portion of time and effort, and had to be done across multiple phases:

1. *Removing tables* – due to the fact that this thesis is focusing on extracting narrative information from annual reports, tables and other figures are not needed in the final summary, as they are considered non-narrative as they mainly include numerical figures and other financial information.
2. *Removing unwanted information* – repeated information such as titles, e-mails, and contact information is not helpful in a summary. As a result, stripping them from the text would allow us to have a better chance of accessing the important useful information within the annual reports through the extractive summarization model.
3. *Eliminating empty lines* – as stated previously, this was a major issue with the dataset, as it was extremely hard to read through the text with random white spacing throughout and eliminating that would allow for easier readability of the generated summaries.

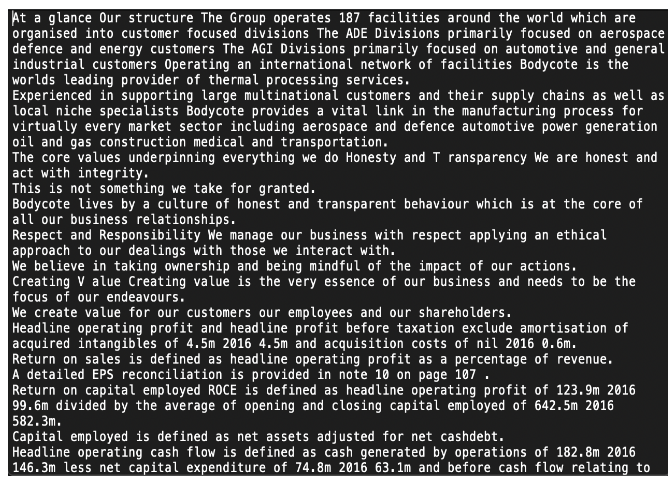
4. *Separating individual sentences* – the unorganized nature of the sentences makes it extremely difficult to efficiently extract important sentences from the text, in order for them to be used in the final summaries, so it is important to isolate each sentence on a separate line.

The steps detailed above were achieved through the use of Regular Expressions to clean the text from unwanted information, while also making use of the pretrained sentence level tokenizer, PunktSentenceTokenizer, found in the NLTK library to tokenize the text at the sentence-level through the use of punctuation and syntactic analysis. and placing the sentences on separate lines.

Figure 1 illustrates an example of text extracted from an annual report before and after preprocessing.



```
 25695    19 March 2018 3:29 PM    Proof 7 25695    19 March 2018 3:29 PM    Proof 7
annual report 2017
Bodycote plc annual report 2017
www.bodycote.com | Stock code: BOY 25695    19 March 2018 3:29 PM    Proof 7 25695    19
March 2018 3:29 PM    Proof 7
At a glance
Our structure
The Group operates 187 facilities around the world which are organised into
customer focused divisions:
The ADE Divisions
(primarily focused on aerospace, defence and energy customers)
The AGI Divisions
(primarily focused on automotive and general industrial customers)
Operating an international network of facilities, Bodycote is the world's
leading provider of thermal processing services.
Experienced in supporting large multinational customers and their supply
chains, as well as local niche specialists, Bodycote provides a vital link in the
manufacturing process for virtually every market sector including aerospace and
defence, automotive, power generation, oil & gas, construction, medical and
transportation.
The core values underpinning everything we do
Honesty and T ransparency
We are honest and act with integrity. This is not something we take for granted. Bodycote
lives by a culture of
honest and transparent behaviour, which is at the core of all our business relationships.
Respect and Responsibility
We manage our business with respect, applying an ethical approach to our dealings with
those we interact with. We
believe in taking ownership, and being mindful of the impact of our actions.
Creating V alue
```



```
At a glance Our structure The Group operates 187 facilities around the world which are
organised into customer focused divisions The ADE Divisions primarily focused on aerospace
defence and energy customers The AGI Divisions primarily focused on automotive and general
industrial customers Operating an international network of facilities Bodycote is the
worlds leading provider of thermal processing services.
Experienced in supporting large multinational customers and their supply chains as well as
local niche specialists Bodycote provides a vital link in the manufacturing process for
virtually every market sector including aerospace and defence automotive power generation
oil and gas construction medical and transportation.
The core values underpinning everything we do Honesty and T ransparency We are honest and
act with integrity.
This is not something we take for granted.
Bodycote lives by a culture of honest and transparent behaviour which is at the core of
all our business relationships.
Respect and Responsibility We manage our business with respect applying an ethical
approach to our dealings with those we interact with.
We believe in taking ownership and being mindful of the impact of our actions.
Creating V alue Creating value is the very essence of our business and needs to be the
focus of our endeavours.
We create value for our customers our employees and our shareholders.
Headline operating profit and headline profit before taxation exclude amortisation of
acquired intangibles of 4.5m 2016 4.5m and acquisition costs of nil 2016 0.6m.
Return on sales is defined as headline operating profit as a percentage of revenue.
A detailed EPS reconciliation is provided in note 10 on page 107 .
Return on capital employed ROCE is defined as headline operating profit of 123.9m 2016
99.6m divided by the average of opening and closing capital employed of 642.5m 2016
582.3m.
Capital employed is defined as net assets adjusted for net cashdebt.
Headline operating cash flow is defined as cash generated by operations of 182.8m 2016
146.3m less net capital expenditure of 74.8m 2016 63.1m and before cash flow relating to
```

Figure 1:a sample of text extracted from an annual report

before and after preprocessing

4. Methodology

In this section, we describe the methodology we explored to build an extractive summarization system for annual reports. We followed two main approaches: Sentence-based extractive summarization and Section-based extractive summarization. Before presenting these approaches, we will start by explaining the NLP technique we followed to encode sentences in a computational format.

4.1. BERT for Text Encoding

Sentence Encoding/Embedding is an upstream task required in our task. The goal is to represent a variable length sentence into a fixed length vector (e.g. hello world to [0.1, 0.3, 0.9]). Each element of the vector should "encode" some semantics of the original sentence.

In order to encode each sentence in the form of a sentence embedding, we made use of the Bidirectional Encoder Representations from Transformers (BERT) langage model (Devlin et al. 2018). BERT is a NLP model developed by Google for pre-training language representations. It leverages an enormous amount of plain text data publicly available on the web and is trained in an unsupervised manner. The model is essentially trained to detect and represent sentences within a text through transformer vector embeddings, by looking at the relationship between words in a text, and processing a whole piece of textual input at once. BERT works by tokenizing each sentence by adding a " [CLS] token … at the beginning of the first sentence [of the input text] and a [SEP] token... at the end of each sentence." (Horev, 2018) Ultimately, this leads to the model being able to form a sentence embedding from the embeddings of the tokenized elements of that sentence.

BERT is basically an Encoder stack of transformer architecture. A transformer architecture is an encoder-decoder network that uses self-attention on the encoder side and attention on the decoder side.. These are more than the Transformer architecture described in the original paper (6 encoder layers). BERT architectures also have larger feedforward-networks, and more attention heads than the Transformer architecture suggested in the original paper. It contains 512 hidden units and 8 attention heads.

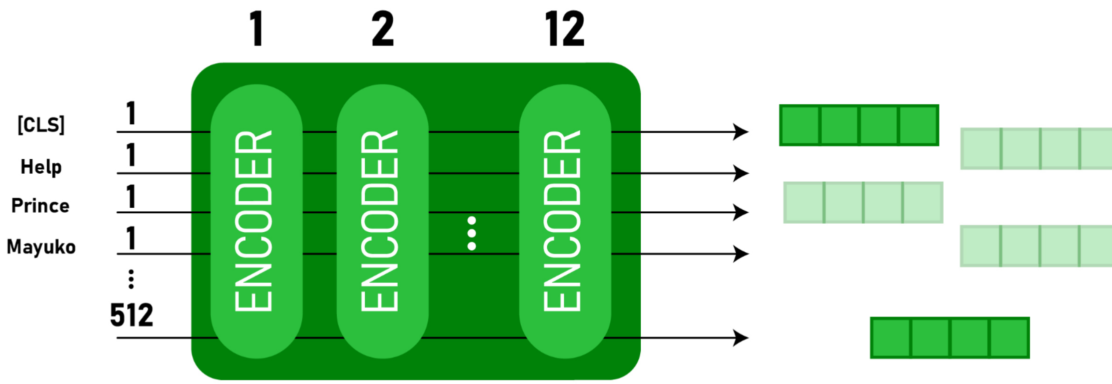


Figure 2 BERT output as Embeddings[2]

BERT Pre-trained Models Explored

Pre-training a BERT model is a fairly expensive yet one-time procedure for each language. Fortunately, **m**any pretrained models were made publicly available. We decided to explore the following three models, due to their relevance to this specific financial text summarization task at hand:

1. *bert-uncased-large*: Most downloaded model overall (15,400,000 downloads) . It is pretrained on BookCorpus: consists of 11,038 unpublished books and English Wikipedia
2. *distilbart-cnn-12-6*: Most downloaded model for summarization (184,000 downloads. This model is pretrained on CNN/DailyMail dataset: consists of over 300,000 unique news articles, and their corresponding highlights (summaries) as written by journalists at CNN and the Daily Mail (Chen, 2017)
3. *FinBert*: Model for sentiment analysis on financial data. This model is pPretrained on Financial PhraseBank: consists of 4845 english sentences selected from financial news that were annotated

[2] Figure taken from: https://www.geeksforgeeks.org/explanation-of-bert-model-nlp/

by 16 people with background in finance and business, as to whether the sentence might negatively or positively affect company stock price. (Malo et al., 2013)

It is important to note that BERT has a maximum input length of 512 tokens at any one given time, meaning that it is unable to compute large amounts of data at once, which is why in this task, we computed individual sentence embeddings, and then computed whole section or document embeddings by taking an average of the relevant sentence embeddings. A BERT model could be trained and fine tuned using different types of data for different contexts. For instance, there are models that are used to identify question & answer textual information, while other models are more geared towards certain dialects and languages. In fact, there are a total of 8498 pretrained models available at the moment, through the HuggingFace transformers library (Wolf et al., 2020).

4.2. Sentence-based Summarization Approaches

As discussed previously, an extractive summarization task typically consists of extracting the most relevant individual sentences from the original document. As a result, we decided to initially approach this task by creating sentence-based summaries, which included extracting individual sentences from the original documents, and only including the most relevant ones in the final summaries.

4.2.1. Sentence-based summarization as a Classification task

As an initial method, we formulated the extractive summarization task as a binary sentence classification task that assigns 1 to a given sentence if it is to be kept in the summary, and 0 if it is to be discarded.

Data annotation: In order to do this, we needed to build a labelled training set. This process involved going through every sentence in all the annual reports in the training dataset, and checking if it exists in the corresponding reference summary. If it exists, it is labelled as 1, otherwise it is labelled a 0. This method is visualized in Figure 3 below.

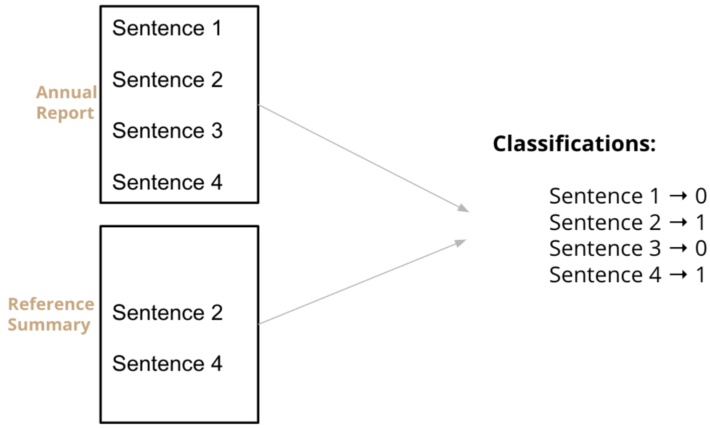


Figure 3: Sentence labeling process for the classification task

Building the Classifier: Once data annotation was done, the appropriate classifier had to be built by finetuning our own BERT model using the annotated data. The data was processed through BERT in order for it to be tokenized appropriately using the Pre-trained BERT Tokenizer. In order to feed the tokenized data into the BERT model, we converted the tokenized data into tensors using PyTorch. With the tensors being created, we were now able to finetune our BERT model by defining our training arguments, and passing the model and those arguments into the trainer, and training the model using the built-in Trainer() class, which 'uses a built-in default function to collate batches and prepare them to be fed into the model" ("Training and Fine-Tuning, n.d.), to train and evaluate any Transformers model.

Running the model on Test Set: After building and training the classifier, we ran the model on the test set to classify the sentences. One major issue with this approach seemed to be the major imbalance in the training dataset. As discussed previously, annual reports are long, and usually contain thousands of lines. Out of these sentences, only 50 to 100 of them would be found in the corresponding reference summaries and labelled as 1, while the thousands of other sentences would be labelled as 0. This means that once the sentences were classified and fed into the BERT encoder, every sentence in the test set would be classified as 0, since the training data had major imbalance in the sentences classified as 0 vs. classified as 1. As a result of this imbalance, we needed to explore another approach that is more tailored to this problem.

4.2.2. CENTROID-SENTENCE-BASED method

Since, as discussed earlier, individual sentences were identified using PunktSentTokenizer, these sentences were inputted into the BERT model, in which they would be tokenized, and a vector embedding representation for each sentence was created. Once those embeddings were created, a 'centroid' vector embedding representation of the whole document was determined by finding the mean of all the sentence embeddings. This approach is shown in Figure 4 below:

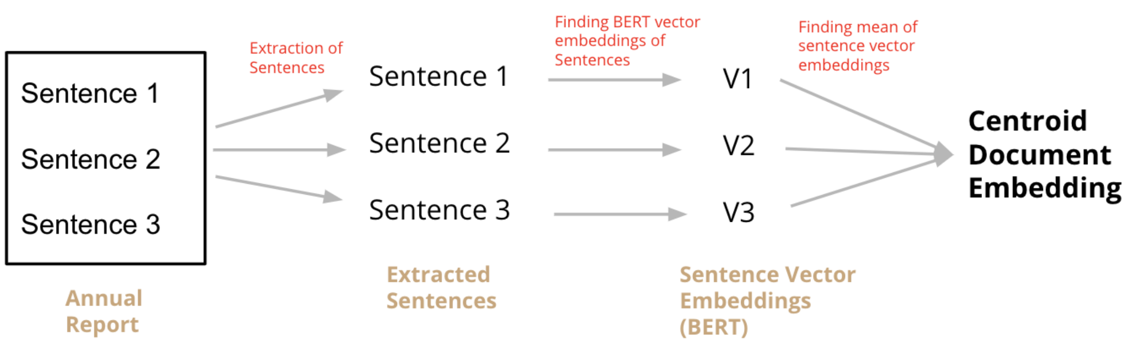


Figure 4: Determining Centroid Document Embedding

To identify the most important sentences in a document, we used Cosine similarity to determine the level of similarity between each individual sentence's embedding and the centroid vector embedding. Initially, the top 30 sentences with the highest similarity scores were kept in the final summary and the rest were discarded. However, it was apparent that the top sentences extracted were not coherent since they were not consecutive within the original document, which made the resulting summaries difficult to comprehend. As a result, a different approach had to be explored, which included extracting the top three sentences, while also extracting the seven sentences surrounding each of the three sentences, which also made for a total of 30 sentences.

Although the resulting summaries seemed to be decent summaries, due to the nature of the task being explored and the available dataset, once the summaries were evaluated, we realized that the reference summaries in the dataset were mainly extracted from whole sections within the original annual reports, rather than individual sentences. Furthermore, creating another dataset that caters to sentence-based summaries would require financial and temporal resources that were not necessarily available to us. This meant that we had to redefine our extractive summarization task by switching from sentence-based summarization to section-based summarization, in order for us to cater to the available dataset.

4.3 Section-based Summarization Approaches

4.3.1 Section Extraction:

As defined by Litvak et al. (2020), there are typically 13 predefined narrative section titles found in an annual report:

> ***sections_types*** *= ['"chairmans statement"', '"chief executive officer ceo review"', '"chief executive officer ceo report"', '"governance statement"', '"remuneration report"', '"business review"', '"financial review"', '"operating review"', '"highlights"', '"auditors report"', '"risk management"', '"chairmans governance introduction"', '"Corporate Social Responsibility CSR disclosures"']*

Using regular expressions, we identified the presence of titles within the annual reports by specifying that they (1) contain less than 5 words (2) could be placed on separate lines (3) are uppercase words. Once those extracted titles were identified, they were compared with the 13 predefined titles using Spacy's similarity API. Extracted titles with a similarity score of 0.7 or higher were considered the chosen titles. Then, text between consequent chosen titles was taken as a whole section.

4.3.2. SECTION-COSINE

Once the task was somewhat redefined, we decided to apply a similar approach like the one previously explored in the sentence-based summarization, which included comparing embeddings of individual sentences with the average embedding of the whole document. However, in this case, rather than comparing the centroid document embedding with sentence embeddings, we determined average vector embeddings for each section by averaging the embeddings of the sentences within a section. Once the cosine similarity score for each section was determined, the first 1000 words of the section with the highest similarity score were extracted from the original document, and included in the final summary.

4.3.3. SECTION-CLUSTERING

To further investigate relevant approaches within the domain, we decided to explore the idea of clustering in extractive summarization. Clustering is typically performed on sentence-based summaries, as you would typically determine the number of clusters you have by equating it to the number of sentences in the final summary. Once the sentences are clustered, “Each cluster of sentence embeddings can be interpreted as a group of semantically identical sentences which more or less carries the same information and whose meaning can be represented by only one sentence from the cluster.” (Gupta, 2020) Then the top sentence from each cluster is chosen by extracting the sentence with the top similarity score when compared to the centroid within its corresponding cluster. In this case, however, rather than clustering individual sentences, whole section embeddings were clustered. The number of clusters in this case was 13 because we used a list of 13 predefined section titles that are usually found within the average annual report. A total of 14,808 sections from the training and validation datasets were clustered. Then, the most common section within each cluster was determined by looking at the frequency of their appearance in the reference summaries of the training and validation datasets, and the average embedding for each document in the testing dataset was compared to each of the 13 centroids (i.e. sections) using cosine similarity, to determine which of the centroids most similar to. The top three closest centroids/sections for each document were then determined, and the first 1000 words of the top existing section in the original document were included in the final summary.

4.3.4. WEIGHTED-SECTION-CLUSTERING

To further build on the clustering approach, rather than just taking the top available section as an accurate summary of the document, we decided to consider the weight of each of the 13 sections by looking at the frequency of their appearance in the reference summaries of the training and validation datasets. To get the weights of each section title, all the reference summaries of the training and validation sets were compared to the extracted sections of their corresponding document, using Spacy’s similarity API. We labelled each reference summary with the name of the section that it was the most similar to. A frequency counter was then created to keep track of all the reference summaries that correspond to a certain section title. In doing so, we determined the weight of each section in the reference summaries, as shown below in Figure 5.

"chairmans statement" 0.257368580978426

"chief executive officer ceo review" 0.2147270333232047

"auditors report" 0.0996657550896384

"highlights" 0.09804517370606705

"financial review" 0.08659981768459435

"business review" 0.07333130760660386

"remuneration report" 0.06857084979236301

"risk management" 0.03352577737263243

"chief executive officer ceo report" 0.033019345690266384

"operating review" 0.01610452749924035

"governance statement" 0.014787805125088626

"chairmans governance introduction" 0.00425402613187481

Figure 5: Weights of each of the predefined narrative section titles

With these frequency weights, once the top three sections for each document were determined through the clustering process described above, rather than simply taking the top existing section, this time the existing section with the highest frequency weight was prioritized, and the top 1000 words of that section were included in the final summary.

5. Evaluation and Results

We evaluated the summaries produced by the methods described in section 4, both intrinsically and extrinsically.

5.1. Intrinsic evaluation

To evaluate the quality of the generated summaries, the common metric in the Automatic Text Summarization field is called ROUGE score. ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It works by comparing an automatically produced summary or translation against a set of reference gold summaries (typically human-produced) (Lin, 2004). Four ROUGE metrics were used, which include: ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-SU4. For ROUGE-1 and ROUGE-2, which both fall under ROUGE-N, they are determined by matching overlap of unigrams and bigrams, respectively, of the generated and reference summary. ROUGE-L is considered the most important metric in summarization, as it "measures sentence-to-sentence similarity based on the longest common subsequence statistics between a candidate translation and a set of reference translations." (Lin & Och, 2004) While ROUGE-SU4 "uses bigrams with a maximum skip distance of 4 between." (Conroy, Schlesinger, & O'Leary, 2011)

Two baseline techniques were used to compare the performance of our systems:

- **TextRank:** An unsupervised text summarization technique inspired by the PageRank algorithm used primarily for ranking web pages in online search results. It is a graph-based ranking that constructs a network of keywords, the main component for measuring similarity between sentences and extracting key ones, and is generally used as a baseline for text summarization (Mihalcea & Tarau, 2004).
- **LexRank:** A graph-based unsupervised technique that relies on sentence connectivity. "In this model, a connectivity matrix based on intra-sentence cosine similarity is used as the adjacency matrix of the graph representation of sentences" (Erkan & Radev, 2004).

The results given in Table 2 compare the sentence-based and section-based systems to the performance of the baseline methods, and the scores of the top three submissions in the FNS-2020 shared task:

MODEL	Rouge-1			Rouge-2			Rouge-L			Rouge-Sum4		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1
Baseline Methods												
TextRank	0.41	0.12	0.17	0.23	0.04	0.07	0.24	0.20	0.21	0.30	0.05	0.08
LexRank	0.34	0.27	0.26	0.19	0.11	0.12	0.21	0.26	0.22	0.25	0.12	0.14
Shared Task Top Submissions												
SRIB2020-3	0.61	0.39	0.47	0.45	0.22	0.29	0.61	0.38	0.46	0.51	0.21	0.29
SUMSUM-BERT	0.45	0.53	0.46	0.37	0.3	0.31	0.3	0.39	0.32	0.41	0.27	0.3
FORTIA-1	0.43	0.43	0.41	0.3	0.28	0.27	0.4	0.4	0.38	0.34	0.33	0.32
Sentence-Based Experiments (using different models)												
CENTROID-Sentence-based-'bert-uncased-large'	0.37	0.21	0.26	0.14	0.06	0.08	0.30	0.19	0.23	0.23	0.08	0.11
CENTROID-Sentence-based-FinBert	0.34	0.17	0.22	0.07	0.05	0.06	0.23	0.17	0.19	0.12	0.09	0.10
CENTROID-Sentence-based-'distilbart-cnn-12-6'	0.34	0.19	0.23	0.07	0.06	0.07	0.25	0.18	0.20	0.14	0.09	0.10
Section-Based Experiments (using 'original-bert-uncased-large' model)												
SECTION-COSINE	0.47	0.27	0.33	0.25	0.12	0.15	0.41	0.26	0.31	0.33	0.13	0.17
SECTION-CLUSTERING	0.46	0.29	0.33	0.24	0.13	0.16	0.40	0.28	0.32	0.32	0.14	0.17
WEIGHTED-SECTION-CLUSTERING	0.48	0.37	0.38	0.30	0.18	0.21	0.45	0.35	0.36	0.36	0.19	0.22

Table 2: Official Results - Averages scores over all gold summaries corresponding to each annual report., compared to baselines and top FNS-2020 submissions.

Alternative Results - Validation Set

Since there are multiple reference summaries corresponding to each annual report, we decided to filter the reference summaries down to the reference summary with the highest score when compared to our generated summaries in the validation set. In doing so, we are more likely to understand the true effectiveness of our multiple approaches, as some reference summaries were extremely unstructured and would penalize the evaluation scores. The scores using this method are shown in Table 3 below:

	Rouge-1			Rouge-2			Rouge-L			Rouge-Sum4		
MODEL	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**	**R**	**P**	**F1**
CENTROID-Sentence-Based	0.41	0.37	0.39	0.10	0.09	0.10	0.40	0.36	0.38	0.16	0.15	0.15
SECTION-COSINE	0.56	0.50	0.53	0.34	0.30	0.32	0.55	0.49	0.52	0.38	0.34	0.36
SECTION-CLUSTERING	0.57	0.50	0.53	0.34	0.29	0.29	0.57	0.50	0.53	0.38	0.33	0.35
WEIGHTED-SECTION-CLUSTERING	**0.64**	**0.60**	**0.62**	**0.49**	**0.44**	**0.46**	**0.63**	**0.60**	**0.62**	**0.51**	**0.47**	**0.50**

Table 3: Alternative Results - Average of scores of top gold summaries conducted on Validation Set

5.2 Extrinsic Evaluation

At the end of the day, these summaries are created for individuals to make use of them in a beneficial way within the field of Finance. As a result, it was important to conduct existric evaluation to understand how to improve upon the generated summaries and methods used in future explorations. Although this thesis was initially focused on the summarization methods and the intrinsic quantitative evaluation of the generated summaries, and we did not necessarily have much time towards the end of this exploration, we still wanted to get a slight understanding about how these summaries perform when given to some people in the field.

5.2.1. Experimental Design

In order to achieve this, we contacted three Business Administration Final-year students from Carnegie Mellon University in Qatar, about to complete their Finance Track, and conducted the following steps:

1. We gave each of them two pairs of system-generated summaries to evaluate: One sentence-based summary and one section-based summary created by the WEIGHTED-SECTION-CLUSTERING system (of the same text)
2. We asked them to rate each of the two summaries out of 10 (in terms of its effectiveness as a summary)

5.2.2. Extrinsic Evaluation Results

	Participant 1	**Participant 2**	**Participant 3**	**Average Effectiveness Score**
Sentence-Based_1	7	8	7	**7.33**
Section-Based_1	6	7	7	**6.67**
Sentence-Based_2	4	5	7	**5.33**
Section-Based_2	4	6	5	**5**

Table 4: Extrinsic Evaluation Quantitative Results

5.3 Discussion

There are multiple observations that could instantaneously be made when looking at the official results in the intrinsic evaluation in Table 2. When comparing the sentence-based and section-based experiments, it is apparent that the section-based attempts perform much better when evaluated using the FNS-2020 shared task dataset. In fact, there seems to be an increase of 0.13 in The Rouge-L F1 Score. This means that section-based methods are much more relevant when exploring extractive summarization using FNS-2020 dataset. Furthermore, when analyzing the scores of the section-based approaches, the best-performing system seems to be the WEIGHTED-SECTION-CLUSTERING. When this system is compared to the two baseline methods, there is a major increase in both recall and precision, which ultimately led to an increase in the Rouge-L F1-score, from 0.21-0.22 to 0.36. When comparing the official results to the alternative results, we see an immediate improvement in scores due to the fact that the systems aren't being penalized for some of the reference summaries that do not match the generated system summaries. Rouge-L F1-Score increases from 0.36 to a much higher score of 0.61. Finally, referencing the top 3 shared task submissions, and specifically comparing the ROUGE-L F1-Scores, we can see that our proposed system, WEIGHTED-SECTION-CLUSTERING, would rank third amongst the top shared-task systems, with a score of 0.36. These shared task submissions had access to much more advanced section extraction tools, which could have positively impacted their scores, which leaves room for improvement in that aspect of this research.

In terms of extrinsic evaluation, we can make very interesting observations when looking at Table 4 that actually show that sentence-based summaries evaluated better for both texts read by the three participants, with average scores of 7.33 and 5.33, when compared with the section-based summary scores of 6.67 and 5.00, respectively. In fact, one of the participants stated that "When [they] compared the two versions, [they] thought that Section-Based_2 included some information that was not necessary to [their] understanding of the annual report. For instance, information like "I am deeply proud to have been the Chairman of this great Company"" In other words, extracting a whole section might mean that a lot of the information within that specific section might not be central enough to the annual report. This indicates that there needs to be further plans on exploring this extractive summarization task through the sentence-based methods, and with the appropriate dataset for evaluation.

6. Conclusion and Future work

During this research, we explored the task of automatic extractive summarization on Financial documents, specifically annual reports. Many different approaches were explored, and they all fall under two categories: sentence-based summarization, which involves extracting the most informative sentences from the annual reports, and section-based summarization, which involves extracting the most informative section of the annual report. Due to the nature of the FNS-2020 dataset being used, section-based approaches seemed to have much higher intrinsic evaluation scores. Specifically, the WEIGHTED-SECTION-CLUSTERING system yields the best results when evaluated against the testing set, achieving a ROUGE-L F1-Score of 0.36. However, the extrinsic evaluation that involved three Business

Administration students specializing in Finance, found that sentence-based summaries were more effective at conveying the main points of an annual report, on average, achieving scores of 7.33 and 5.33, while section-based summaries of the same texts scored of 6.67 and 5.00, respectively.

In the future, many limitations of this research could be addressed and improved through:

1. *Section Extraction Tools:* one important component to improve in this research is finding a more accurate way to identify and separate the different narrative sections within an annual report. This could be through collaborating with larger entities in the NLP industry with advanced systems in section identification that could help improve the performance of the automatic summarization system.
2. *Creating sentence-based dataset:* As observed in the extrinsic evaluation, sentence-based summaries seemed to read better, on average, when compared to their section-based alternatives. As a result, there should be some focus on exploring the task of generating sentence-based summaries. However, in order to do so, creating a large dataset of annual reports along with their corresponding reference sentence-based summaries is a crucial step in ensuring that the task is explored and evaluated appropriately.
3. *Extensive Extrinsic Evaluation:* Since we conducted a small extrinsic evaluation at the end of this research due to time constraints, it is important to further elaborate on that by including a larger number of human subjects. This could include asking a whole class of university students specializing in Finance to provide further qualitative and quantitative results.
4. *Theme-based summarization:* An interesting exploration could be looking into creating a theme-based summarizer that looks at extracting a specific type of information from the annual reports, such as the future plans of the organization. This gives potential investors a chance to evaluate a company's performance through identifying and analyzing their future plans.

References

Abujar, S., Hasan, M., Shahin, M. S. I., & Hossain, S. A. (2017). A heuristic approach of text summarization for Bengali documentation. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. Retrieved from http://dx.doi.org/10.1109/icccnt.2017.8204166

Azzi, A. A., Bouamor, H., & Ferradans, S. (2019). The FinSBD-2019 Shared Task: Sentence Boundary Detection in PDF Noisy Text in the Financial Domain. *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, 74–80. Retrieved from https://www.aclweb.org/anthology/W19-5512

Azzi, A. A., & Kang, J. (2020). FNS-Summarisation 2020 shared task: system description paper Extractive Summarization System for Annual Reports. *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*. Presented at the Barcelona, Spain.

Berger, A., & Mittal, V. O. (2000). Query-relevant summarization using FAQs. *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics - ACL '00*. Retrieved from http://dx.doi.org/10.3115/1075218.1075256

Chen, J.-H. (2009). Time value of money and its applications in corporate finance: A technical note on linking relationships between formulas. *American Journal of Business Education (AJBE)*, *2*(6), 77–88. https://doi.org/10.19030/ajbe.v2i6.4090

Chen, V. (2017). An Examination of the CNN / DailyMail Neural Summarization Task.

Chuang, W. T., & Yang, J. (2000). Extracting sentence segments for text summarization. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '00*. Retrieved from http://dx.doi.org/10.1145/345508.345566

Conroy, J. M., Schlesinger, J. D., & O'Leary, D. P. (2011). Nouveau-ROUGE: A novelty metric for update summarization. *Computational Linguistics*, *37*(1), 1–8.

https://doi.org/10.1162/coli_a_00033

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*. Retrieved from http://dx.doi.org/10.18653/v1/n19-1423

El-Haj, M., Dr., Rayson, P., & Moore, A. (2018). Towards a Multilingual Financial Narrative Processing System. *The First Financial Narrative Processing Workshop: Proceedings of the 11th Edition of the Language Resources and Evaluation Conference*, 52–58.

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, *22*, 457–479. https://doi.org/10.1613/jair.1523

EY. (2020a). Global Fintech Adoption Index 2019. Retrieved from EY website: https://fintechauscensus.ey.com

Filippova, K., Surdeanu, M., Ciaramita, M., & Zaragoza, H. (2009). Company-oriented extractive summarization of financial news. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics on - EACL '09*. Retrieved from http://dx.doi.org/10.3115/1609067.1609094

Fowler, D. S. (2018). How many websites are there in the world? | tek eye. Retrieved April 17, 2021, from https://tekeye.uk/computing/how-many-websites-are-there

Ghazali, Mohd., & Anum, N. (2010). The importance and usefulness of corporate annual reports in malaysia. *Gadjah Mada International Journal of Business*, *12*(1), 31. https://doi.org/10.22146/gamaijb.5516

Gupta, A. (2020, September 29). Understanding Text Summarization using K-means Clustering. Retrieved February 18, 2021, from Medium website: https://medium.com/@akankshagupta371/understanding-text-summarization-using-k-means-clustering-6487d5d37255

Gupta, V., & Lehal, G. S. (2010). A survey of text summarization extractive techniques. *Journal of*

Emerging Technologies in Web Intelligence, *2*(3). https://doi.org/10.4304/jetwi.2.3.258-268

Horev, R. (2018, November 17). BERT Explained: State of the art language model for NLP. *Towards Data Science*. Retrieved from https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

Hsu, W.-T., Lin, C.-K., Lee, M.-Y., Min, K., Tang, J., & Sun, M. (2018). A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Retrieved from http://dx.doi.org/10.18653/v1/p18-1013

J.P. Morgan. (2020b). Innovation with machine learning. Retrieved April 7, 2021, from https://www.jpmorgan.com/global/research/machine-learning

Khant, A., & Singh Mehta, M. (2018). Analysis of Financial News Using Natural Language Processing and Artificial Intelligence. *Conference: International Conference on Business Innovation 2018*.

Kunder, M. de. (2018). WorldWideWebSize.com | The size of the World Wide Web (The Internet). Retrieved from https://www.worldwidewebsize.com

Leidner, J. L. (2019). Summarization in the Financial and Regulatory Domain. In *Trends and Applications of Text Summarization Techniques* (pp. 187–215). IGI Global.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *Conference: In Proceedings of the Workshop on Text Summarization Branches Out*.

Lin, C.-Y., & Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics - ACL '04*. Retrieved from http://dx.doi.org/10.3115/1218955.1219032

Litvak, M., Vanetik, N., & Puchinsky, T. (2018). Hierarchical summarization of financial reports with RUNNER. *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, 212–224. Barcelona, Spain.

Luo, J., Li, X., & Chen, H. (2018). Annual report readability and corporate agency costs. *China Journal of Accounting Research*, *11*(3), 187–212. https://doi.org/10.1016/j.cjar.2018.04.001

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2013). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, *65*(4), 782–796. https://doi.org/10.1002/asi.23062

Mihalcea, R. (2004). Graph-based ranking algorithms for sentence extraction, applied to text summarization. *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions* -. Retrieved from http://dx.doi.org/10.3115/1219044.1219064

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing Order into Text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404–411. Barcelona, Spain: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/W04-3252

Moratanch, N., & Chitrakala, S. (2016). A survey on abstractive text summarization. *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. Retrieved from http://dx.doi.org/10.1109/iccpct.2016.7530193

Ratia, T. (2020). 20 Applications Of Automatic Summarization In The Enterprise. Retrieved from The Frase Blog website: https:// https://blog.frase.io/20-applications-of-automatic-summarization-in-the-enterprise

Shirwandkar, N. S., & Kulkarni, S. (2018). Extractive text summarization using deep learning. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. Retrieved from http://dx.doi.org/10.1109/iccubea.2018.8697465

Singh, A. (2020). PoinT-5: Pointer Network and T-5 based Financial Narrative Summarisation. *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, 104–110.

Stanton, P., & Stanton, J. (2002). Corporate annual reports: Research perspectives used. *Accounting, Auditing & Accountability Journal*, *15*(4), 478–500. https://doi.org/10.1108/09513570210440568

Training and fine-tuning. (n.d.). Retrieved April 25, 2021, from HuggingFace website: https://huggingface.co/transformers/training.html

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., … Rush, A. (2020). Transformers: State-of-the-Art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Retrieved from http://dx.doi.org/10.18653/v1/2020.emnlp-demos.6

Xu, W., Grishman, R., Meyers, A., & Ritter, A. (2013). A Preliminary Study of Tweet Summarization using Information Extraction. *Proceedings of the Workshop on Language in Social Media (LASM 2013)*, 20–29.

Yang, J., Parekh, R., & Honavar, V. (1999). DistAl: An inter-pattern distance-based constructive learning algorithm. *Intelligent Data Analysis*, *3*(1), 55–73. https://doi.org/10.3233/ida-1999-3105

Zheng, S., Lu, A., & Cardie, C. (2020). SUMSUM@FNS-2020 Shared Task. *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, 147–151.