

# DiMSum: Distributed and Multilingual Summarization of Financial Narratives

Neelesh K Shukla, Amit Vaid, Raghu C Katikeri, Sangeeth Keeriyadath, Msp Raja

State Street Corporation

Bengaluru, India

{nshukla, avaid, rkateri, skeeriyadath, smaddila1}@statestreet.com

## Abstract

This paper was submitted for Financial Narrative Summarization (FNS) task in FNP-2022 workshop. The objective of the task was to generate not more than 1000 words summaries for the annual financial reports written in English, Spanish and Greek languages. The central idea of this paper is to demonstrate automatic ways of identifying key narrative sections and their contributions towards generating summaries of financial reports. We have observed a few limitations in the previous works: First, the complete report was being considered for summary generation instead of key narrative sections. Second, many of the works followed manual or heuristic-based techniques to identify narrative sections. Third, sentences from key narrative sections were abruptly dropped to limit the summary to the desired length. To overcome these shortcomings, we introduced a novel approach to automatically learn key narrative sections and their weighted contributions to the reports. Since the summaries may come from various parts of the reports, the summary generation process was distributed amongst the key narrative sections based on the weights identified, later combined to have an overall summary. We also showcased that our approach is adaptive to various report formats and languages.

**Keywords:** distributed, financial, narrative, summarization, multilingual

## 1. Introduction

With increased liberalization across the globe and the sprawling of organizations competing in multiple and varied overlapping sectors, a holistic comparison, and contrast of annual financial reports are in greater demand. Experts are looking for a concise and precise summary of an organization's financial health and future direction to gauge their investment and strategic positions. With the increasing volume of available financial information, the study of NLP methods that automatically summarize the content has grown rapidly into a major research area. A series of Financial Narrative Processing workshops (El-Haj et al., 2022; El-Haj et al., 2021; El-Haj et al., 2020) focused on this area and have invited researchers to participate. The Financial Narrative Summarization (FNS-2022) task (Zmandar et al., 2022) aims to demonstrate the value and challenges of applying automatic text summarization to financial reports written in different languages: English, Spanish and Greek. Financial reports are a bit more challenging than news articles, because companies usually produce glossy brochures with a much looser structure, they are large, contain financial statements and vast information which deemed repetitive. Instead of summarizing the complete report, the task requires locating key narrative sections found in the annual reports and generate a single structured summary for them in not more than 1000 words. "Narrative sections" or "front-end" sections usually contain textual information and reviews by the firm's management and board of directors. Sections containing financial statements in terms of tables and numbers are usually referred to as "back-end"

sections and are not supposed to be part of the narrative summaries. The task dataset has been extracted from annual reports published in PDF file format. These extracted reports were very noisy, making the task even more challenging.

Previous participating systems used a variety of approaches but have one of these limitations. Generating summaries from the complete report instead of identifying narrative sections to summarize or relying on language summarizers to automatically identify the salient sentences and areas without using the contexts of narrative sections. (Litvak and Vanetik, 2021; Krimberg et al., 2021), using heuristics or manual investigations to identify the narrative sections (Orzhenovskii, 2021; Gokhan et al., 2021; Li et al., 2020). A very important aspect of summarization is to produce a short and clear summary with the limits of words or sentences. But while generating a final summary of K words, most of the approaches didn't pay much attention and have lost some part of the novelty in this process, either by taking top K words (Litvak and Vanetik, 2021; Orzhenovskii, 2021) or by dropping sections completely (Ait Azzi and Kang, 2020) or treating all sections equally (Litvak et al., 2020).

We approached the problem by focusing on two sub problems: 1) Automatically identify the key narrative sections (in English reports) or narrative areas (in Spanish/Greek reports), from where the summary needs to be generated, 2) Quantify the contributions of these key narrative sections or areas towards summary in terms of number of words to be generated. To the best of our knowledge, this is the first time, that the

distribution of words has been explored. These can now be fed to a summarizer to generate summaries from individual narrative sections in a distributed manner to be combined later for an overall K-words summary.

## 2. Dataset

FNS-2022 dataset contains annual reports produced by UK, Spanish and Greek firms listed on stock exchange market of each of those countries. English dataset was randomly split into training (75%), testing and validation (25%). This is a bit different for Greek and Spanish as we have a smaller dataset, the split for each language is training (60%), testing and validation (40%). Experts have considered few relevant sections from the annual reports to generate respective gold standard summaries. On average there are at least 3 gold-standard summaries for each English annual report and 2 gold-standard summaries for Spanish and Greek reports. Table 1 2 and 3 details the split of dataset for all the three languages. We further analyzed these datasets and have these findings:

- Texts extracted from the PDF reports had lot of noise: special characters, unexpected spaces, sentence broken into multiple lines and varied character casing of section headers. While this was mostly the case for English and Greek, the Spanish reports had a much cleaner text.
- Gold summaries for the English training dataset were extracted directly from the reports and had a good overlap while very less overlap was found in Spanish and Greek datasets.
- Almost all of the English training dataset (99.996%) reports were structured with the table of contents (TOCs) and the respective headers provided for each section in the body of the report. This arrangement helped us understand the narrative sections of the report and use them for modeling purposes.
- The Spanish and Greek reports did not have any reliable TOCs or section headers.

Type	Training	Validation	Testing
Report Full Text	3000	363	500
Gold Summaries	9873	1250	1673

Table 1: English Dataset

## 3. Approach

A fundamental problem to solve in summarization is to identify the relevant aspects like sections, paragraphs or sentences and produce them in short and clear format

Type	Training	Validation	Testing
Report Full Text	162	50	50
Gold Summaries	324	100	100

Table 2: Spanish Dataset

Type	Training	Validation	Testing
Report Full Text	162	50	50
Gold Summaries	324	100	100

Table 3: Greek Dataset

with limits on the number of words or sentences. Our approach was focused on addressing these problems considering the financial context presented in these reports by A: Identifying key narrative sections or areas and their respective weights (Section 3.1), and B: Quantifying the contribution of key narrative sections or areas in 'number of words' to be extracted based on the weights (Section 3.2). Later, we pass identified key narrative sections and respective number of words to a summarizer for extracting distributed summaries, later to be combined for an overall summary. We explored various summarizers and techniques to generate and combine summaries as described in Section 3.3

### 3.1. Identifying Key Narrative Sections or Areas with Weights

This section describes our approach of identifying key areas in the reports and their respective weights on datasets based on the formats as detailed in subsequent sub-sections.

#### 3.1.1. Key Narrative Sections and Weights in English Reports

In the English dataset, the presence of TOCs in the reports and section names in the respective gold summaries, we defined narrative section identification as a classification problem, where section can be narrative ('true') or non-narrative ('false').

**Building Annotated Dataset:** To train a section classification model, we built an annotated dataset (Figure 1). For each section in a report a row was created with attributes like section name, section page number, section body length i.e. the number of words. A section was labeled as 'true', if the title was narrative (a title has been considered narrative if it was present in any one of the respective gold summaries) and 'false' otherwise. We applied automatic lookup of section names in the respective gold summaries. This process was repeated for each report in the training dataset.

Further the section title names and page numbers were extracted by parsing the TOCs present in the reports. For parsing TOC, we utilized the methods by (Zheng et al., 2020). Their TOC parsing approach captures

the section names along with the respective start page numbers. Having those page numbers helped us extract the complete sections from the report by extracting the pages from start page number of current section till one page before the next section’s start page number.

**Label Correction:** After annotation, we identified that for many of the sections, the labels were overlapping, marking them both narrative and non-narrative (Table 4). For each unique title, label was corrected to the majority label if the percentage of majority label was above 70% (based on our empirical studies and which also holds true for most frequent sections (Table 4)). Final dataset had total 67893 sections with around 20% of sections labeled as narrative or ‘true’.

Section Title	#Positive	#Negative
board of directors	367 (22%)	<b>1342 (78%)</b>
chairmans statement	<b>1729 (72%)</b>	668 (28%)
chief executives review	<b>811 (70%)</b>	345 (30%)
consolidated balance sheet	152 (13%)	<b>1012 (87%)</b>
consolidated cash flow statement	132 (13%)	<b>872 (87%)</b>
highlights	<b>713 (75%)</b>	240 (25%)

Table 4: Label Distribution in Annotated Dataset Before Label Correction

**Model Training:** Before training the model, the text was processed (removed extra spaces, special characters and punctuation, converted to lower case, performed lemmatization and stemming). Stratified sampling was applied to handle imbalance in the labels while splitting. We experimented with many models and found L2 regularized Logistic Regression to the best performing one with 5-fold cross validation accuracy of 93% with weighted average F1 (.92). F1 scores for ‘true’ and ‘false’ classes were 0.78 and 0.96 respectively.

**Key Narrative Sections and Weights:** Approach for identifying key narrative sections and their weights is shown in Figure 2. Given an English report, TOC was parsed to extract section features: section name, page number, length. With these features, classification model was used to categorize the section as narrative (‘true’) or non-narrative (‘false’). Weight of a section can be defined as probability of it being narrative, assigned by the model.

$$W_i : Pr(narrative = true)$$

### 3.1.2. Key Narrative Areas and Weights in Spanish/Greek Reports

Upon investigating the Spanish reports, we didn’t find the concept of TOC and sections like in English

reports. Though we found TOCs in Greek reports but TOC parsing methods used for English reports were not applicable on the Greek reports. Instead of reinventing the wheel again, we focused on identifying a cluster of sentences defined as ‘Key Narrative Areas’ by adopting our work on the English dataset to other languages based on the following assumptions:

**Assumption 1: Language Independence of Narratives:** The key narratives should be independent of language given all are financial reports. i.e. if ‘Chairman’s Statement’ is a key narrative in English reports, so ‘Declaración del Presidente’ should be in Spanish.

**Assumption 2: Structure Independence:** If narratives are not defined as sections, the presence of narrative keywords or key phrases in a sentence indicates it being part of some narrative.

**Assumption 3: Neighbourhoods Assumption:** If a sentence is part of some narrative, most likely its N neighbouring sentences are also part of the same narrative, defining a set of sentences or paragraph as key narrative area

Given these assumptions, we came up with an approach as depicted in Figure 3 and detailed below:

- Extract top M key narrative section titles from English dataset according to their weights as defined in Section 3.1.1.
- Translate key narrative sections to Spanish and Greek. We used Google Translator API<sup>1</sup> for the same.
- Process and tokenize the translated narrative titles into weighted ‘Narrative Keywords’<sup>2</sup>. Weight of a narrative keyword can be defined as:

$$Wt(w) = \sum Wt(Ns) : w \in Ns$$

where Wt(Ns): Weight of narrative section title Ns.

- Tokenize the report into sentences, process them and compute the weights of sentences based on presence of these narrative keywords as defined below:

$$Wt(S) = \sum Wt(w) : w \in S$$

where Wt(w): Weight of narrative keyword w.

<sup>1</sup><https://pypi.org/project/deep-translator/>

<sup>2</sup>We used nltk (<https://www.nltk.org/>) to process the text and tokenize.

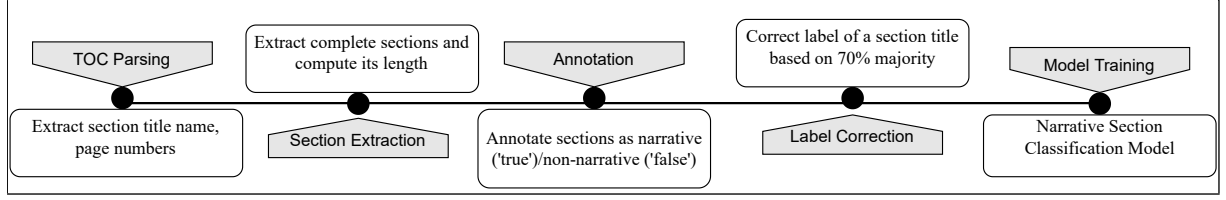


Figure 1: Pipeline for Building Annotated Dataset and Training

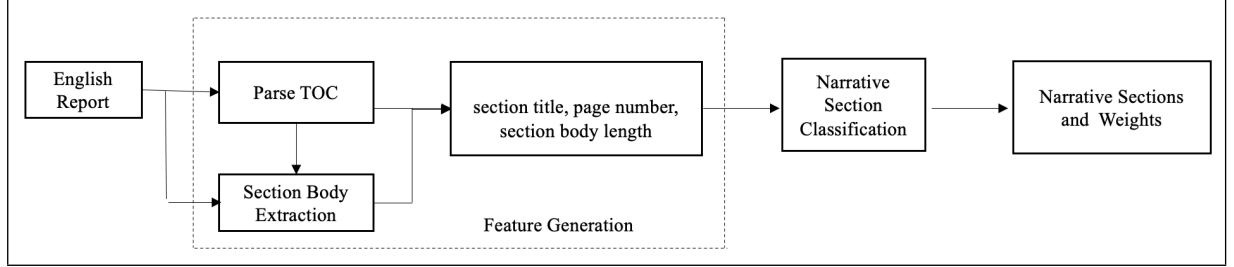


Figure 2: Identifying Key Narrative Sections and Weights in English Reports

- Select top Q sentences (by weight) and its position in original report. These sentences can be assumed as centroid or seed sentences around which key narrative areas can be built.
- For a sentence  $S_i$  at position 'i', key narrative area can be defined as set of sentences from position 'i-N to i+N' as applicable. The weight of this key narrative area can be defined as sum of weights of all sentences in the identified key narrative area.

Key Narrative Area:

$$[S(i - N), \dots, S_i, \dots, S(i + N)]$$

Weight of Narrative Area:

$$\sum Wt(S_j) : S_j \in [S(i - N), \dots, S_i, \dots, S(i + N)]$$

We maintained both raw and processed sentences and summaries were extracted from raw sentences based on position indexes.

Parameters M, Q and N can be fine-tuned for individual dataset. We have fine-tuned to M=50, N=20, Q=2 on respective validation dataset of Spanish and Greek languages.

### 3.2. Quantify the Contribution of Key Narrative Sections or Areas

The goal of summarization system is to generate a brief version of the document that highlights the most salient aspects in a limit on amount of words or sentences as K. In a financial report or in any document, these salient aspects are spread across document with varied subjectivity of being considered for summary. When we looked into gold summaries, we discovered that summaries were coming from various parts of the report.

Based on this observation, we decided to distribute K words among key narrative sections by their respective weights. Sometimes sections do not have enough words in their body as required by the weights assigned, failing to generate complete K word summary, decreasing recall or precision or both. To overcome this problem, we have devised an algorithm called 'K-Maximal Word Allocation' which maximally distributes the required K words among section according to their weights and number of available words in the sections (Algorithm 1). Let's take an example as shown in Table 5. Assume, there are three sections 'section a', 'section b' and 'section c' with their respective weights of 0.9, 0.9 and 0.6. The required number of words for the summary is 1000. In iteration 1, these weights are normalized to the 0-1 scale as 0.375, 0.375 and 0.25. By multiplying 1000 to these weights we can get the number of words required from these sections as 375, 375 and 250. Assume that available numbers of words in respective sections are 75, 500 and 300. With this 'section a' can't generate required 375 words, falling short of 300 words. At the same time other sections 'section b' and 'section c' have extra words 125 (500-275), 50 (300-250) respectively. In iteration 2, we will consider remaining 300 words to be generated for summary, and distribute them in 'section b' and 'section c' according to their new normalized weights. Considering only 'section b' and 'section c', there new normalized weight will be 0.5 and 0.5. These iterations will continue till expected K=1000 have been allocated or number of words in all sections have been exhausted.

### 3.3. Distributed Summary Generation

In previous Sections 3.1 and 3.2, we have identified set of pairs (narrative\_section, num\_words\_to\_be\_generated). Given these inputs, any type and combination of summarizers can be

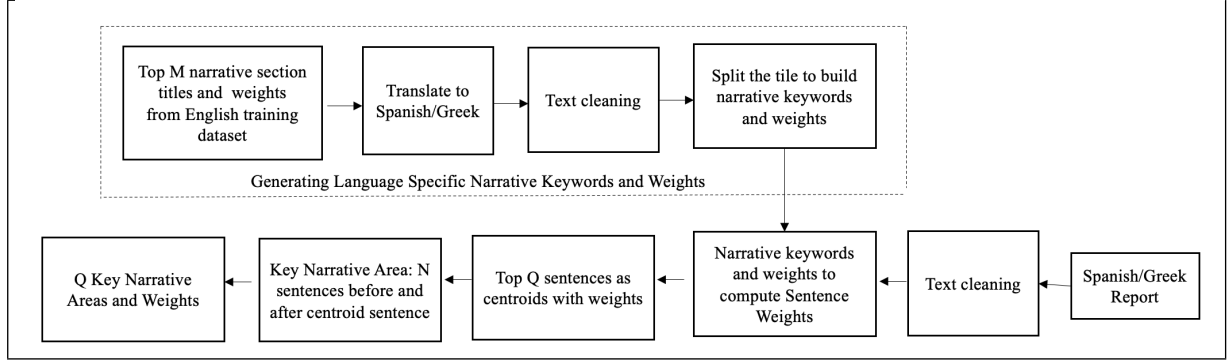


Figure 3: Identifying Key Narrative Areas and Weights in Spanish/Greek Reports

iter.	section	weight (norm weight)	required #words for summary	#words in section	remaining #words required for summary	remaining #words in section
1	section a	0.90 (0.375)	375	75	300	0
1	section b	0.90 (0.375)	375	500	0	125
1	section c	0.60 (0.25)	250	300	0	50

Iteration 1: Required: 1000, Allocated: 700, Remaining Required: 300, Available in Sections: 175

iter.	section	weight (norm weight)	required #words for summary	#words in section	remaining #words required for summary	remaining #words in section
2	section b	0.90 (0.60)	180	125	55	0
2	section c	0.60 (0.40)	120	50	70	0

Iteration 2: Required: 1000, Allocated: 875, Remaining Required: 125, Available in Sections: 0

Table 5: Example of Maximal Word Allocation for 1000-words Summary

used to generate summary as depicted in Figure 4. Each pair is passed to a summarizer to generate a sub summary later to be combined for an overall summary. Various combination approaches can be followed. To have a similar flow as the report, we structured the narrative summaries in order of their respective section’s positions in the original report.

## 4. Experiments and Results

We used ROUGE (Lin, 2004) metrics, ROUGE-1 and ROUGE-2 and evaluated methods on the validation dataset using python package <sup>3</sup>. Since there were multiple golden summaries, for each report, we computed the ROUGE scores with each corresponding summary and took an average.

### 4.1. Comparison of Summarizers

As described in Section 3.3, any summarizer can be used in the distributed summary generation process,

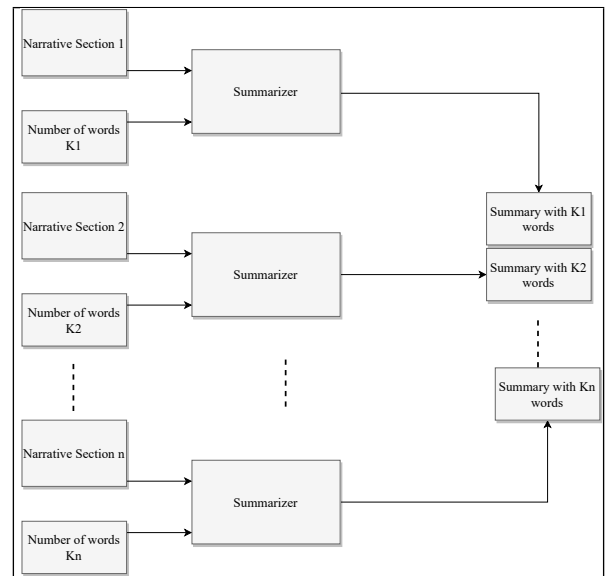


Figure 4: Distributed Summary Generation

<sup>3</sup><https://pypi.org/project/rouge-score/>

---

**Algorithm 1** K-Maximal Word Allocation

---

**Inputs:**

$S_w \leftarrow \text{list\_of\_section\_weights}$   
 $W \leftarrow \text{list\_of\_number\_of\_words\_in\_each\_section}$   
 $K \leftarrow \text{required\_number\_of\_words\_in\_final\_summary}$   
 $K_{Alloc} \leftarrow \text{list\_of\_allocated\_number\_of\_words\_to\_each\_section\_till\_previous\_iterations}$

**procedure** ALLOCATE\_MAXIMAL\_WORDS

```
if  $K = 0$  or  $\text{sum\_of}(S_w) = 0$  then
    return  $K_{Alloc}$ 
end if
 $S_{w\_normalized} = S_w / \text{sum\_of}(S_w)$ 
 $W_{Req} = K \times S_{w\_normalized}$ 
if  $W_{Req} \leq W$  then
    return  $K_{Alloc} + W_{Req}$ 
else
    return  $K_{Alloc} + W$ 
end if
for  $i = 0$  to  $\text{length\_of}(S_w)$  do
    if  $W_{Req}[i] \geq W[i]$  then
         $K_{Alloc}[i] = K_{Alloc}[i] + W[i]$ 
         $K = K - W[i]$ 
         $W[i] = 0$ 
    else
         $K_{Alloc}[i] = K_{Alloc}[i] + W_{Req}[i]$ 
         $K = K - W_{Req}[i]$ 
         $W[i] = W[i] - W_{Req}[i]$ 
    end if
end for
allocate_maximal_words( $S_w, W, K, K_{Alloc}$ )
end procedure
```

---

we compared three extractive summarizers: 1) Top-k summarizer, which extracts the first k words from given text, 2) Google BERT (Devlin et al., 2018) based Extractive Summarizer<sup>4</sup> by (Miller, 2019) and 3) Facebook BART (Lewis et al., 2020) based extractive summarizer provided by Hugging Face<sup>5</sup>. Table 6 shows the results on the English dataset where Top-k summarizer outperformed other summarizers. We used the Top-k extractor for further experiments.

## 4.2. K-Maximal Allocation and Distributed Summary Generation

We built systems using our novel approaches, K-Maximal Word Allocation and Distributed Summary Generation (Sections 3.2, 3.3) on English, Spanish and Greek datasets as *SSC-AI-RG-English*, *SSC-AI-RG-Spanish* and *SSC-AI-RG-Greek* respectively. We used one of the official FNS-2021<sup>6</sup> base-

lines, TextRank (Mihalcea and Tarau, 2004)<sup>7</sup>. As shown in Table 9 our system performed extremely well on English dataset and decently better on Spanish (Table 7) and Greek datasets (Table 8) compared to the baseline. This system was submitted as *SSC-AI-RG-3*.

## 4.3. Alternate Summary Generation on English Dataset

Since complete sections were extracted for gold summaries, we also experimented with alternate summary generation for English dataset. Once the key narrative sections were identified with weights as described in 3.1.1, instead of applying our novel approaches, we extracted complete sections and combined them in, i) ascending order of page number or position in the report (System *SSC-AI-RG-Alt1-English*) and, ii) descending order of weights learned (System *SSC-AI-RG-Alt2-English*). Top-1000 words were extracted to generate summary. These two systems were combined with the Spanish and Greek systems described in Section 4.2, and were submitted as *SSC-AI-RG-1* and *SSC-AI-RG-2* respectively.

*SSC-AI-RG-Alt1-English*, was the best performing one (Table 9). It was due to the nature of the dataset where the majority of the summaries were in Top 10% (Zheng et al., 2020). It can also be observed that our novel summarization approach, *SSC-AI-RG-English* worked pretty well without considering this dataset specific characteristic, showcasing the generic nature of it.

## 4.4. Official Results

The official results are shown in Table 10. Teams were ranked according to ROUGE-2 F1 score on test dataset. With overall score combined across languages, our two systems *SSC-AI-RG-1* and *SSC-AI-RG-3* were in Top-3. Our systems performed best on the Greek dataset and second best on the Spanish one. This demonstrates the effectiveness of our approach in multilingual setup.

## 5. Related Works

(Ait Azzi and Kang, 2020) also defined the problem of narrative section identification as a binary classification system. We would like to highlight a few differences: 1) Our system additionally considers position and length of the section along with its title. 2) Our label correction strategy considers a label change to the majority label only when the proportion exceeds 70%. 3) Compared to their approach of extracting top 1000 words from one section as a summary, we added novelty of generating distributed summary using 'K-Maximal Word Allocation' algorithm as described in Sections 3.2 3.3. Our system achieved better classification accuracy 93% compared to their 70%.

---

<sup>4</sup><https://pypi.org/project/bert-extractive-summarizer>

<sup>5</sup><https://huggingface.co/facebook/bart-large-cnn>

<sup>6</sup>[www.lancaster.ac.uk/staff/elhaj/docs/fns2021\\_results.pdf](http://www.lancaster.ac.uk/staff/elhaj/docs/fns2021_results.pdf)

---

<sup>7</sup><https://pypi.org/project/summa/>

Summarizer	R1P	R1R	R1F	R2P	R2R	R2F
BART	0.544	0.444	0.417	0.304	0.244	0.232
BERT	0.56	0.40	0.42	0.32	0.20	0.22
Top-k	0.523	0.596	<b>0.508</b>	0.347	0.418	<b>0.345</b>

Table 6: Comparison of Summarizers for Generating Distributed Summary on English Validation Dataset

Dataset	R1P	R1R	R1F	R2P	R2R	R2F
SSC_AI_RG_Spanish	0.357	0.566	<b>0.41</b>	0.122	0.192	<b>0.139</b>
TextRank (Baseline)	0.34	0.543	0.393	0.104	0.166	0.12

Table 7: Result on Spanish Validation Dataset

Dataset	R1P	R1R	R1F	R2P	R2R	R2F
SSC_AI_RG_Greek	0.349	0.429	0.385	0.155	0.194	<b>0.172</b>
TextRank (Baseline)	0.532	0.255	0.396	0.259	0.112	0.156

Table 8: Result on Greek Validation Dataset

System	R1P	R1R	R1F	R2P	R2R	R2F
SSC_AI_RG_English	0.523	0.596	<b>0.508</b>	0.347	0.418	<b>0.345</b>
SSC_AI_RG_Alt1_English	0.551	0.643	<b>0.546</b>	0.415	0.512	<b>0.425</b>
SSC_AI_RG_Alt2_English	0.499	0.541	0.478	0.297	0.313	0.281
TextRank (Baseline)	0.321	0.339	0.284	0.084	0.087	0.071

Table 9: Results of Different Systems on English Validation Dataset

Team	English	Spanish	Greek	Overall Score
LSIR-1	0.365	<b>0.157</b>	0.141	0.257
<b>SSC-AI-RG-1</b>	0.327	<b>0.146</b>	<b>0.185</b>	0.24625
<b>SSC-AI-RG-3</b>	0.319	<b>0.146</b>	<b>0.185</b>	0.24225
IIC	0.366	0.125	0.095	0.238
<b>SSC-AI-RG-2</b>	0.3	<b>0.146</b>	<b>0.185</b>	0.23275
Team-Tredence-2	0.322	0.131	0.138	0.22825
Team-Tredence-1	0.317	0.131	0.138	0.22575
LIPI	0.374	0.07	0.046	0.216
Team-Tredence-3	0.322	0.131	0.072	0.21175
LSIR-3	0.275	0.138	0.13	0.2045
MACQUARIE-1	0.303	0	0	0.1515
MACQUARIE-3	0.302	0	0	0.151
MACQUARIE-2	0.301	0	0	0.1505
AO-LANCS	0.143	0.134	0.131	0.13775

Table 10: Official FNS-2022 Results on Test Dataset. Ranked According to ROUGE-2 F1 Score. Overall Score: English (50%), Spanish (25%) and Greek (25%)

(Zheng et al., 2020) also built the classification system. They extracted the first 5 sections, and labeled one section as positive with maximum overlap with gold summaries and others as negative. Whereas we consider all the sections and mark the sections positive if they are present in any of the gold summaries otherwise negative. They took into account the complete section (title+body) for classification whereas we used the titles.

## 6. Conclusion and Future Work

We explored the aspect of finding narrative sections, quantifying their contributions as weights and words to be extracted based on these weights. We introduced a concept of 'Maximal Word Allocation in Summarization' which can be used across problems and domains not limited to financial reports. We also introduced a generic approach that can be adapted to dif-

ferent languages and report formats. In this work, we focused on the inputs and outputs of summarizers. In future work, we would like to explore, we would like to explore more sophisticated approaches for summarization using the foundations that we laid using K-Maximally Allocated Words and Distributed Summary Generations. These concepts are generic enough to be used in any domain with any summarizer. Our current approach is also dependent upon the TOC in English Reports. Alternate approaches need to be explored to reduce this dependency.

## 7. Bibliographical References

- Ait Azzi, A. and Kang, J. (2020). Extractive summarization system for annual reports. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 143–147, Barcelona, Spain (Online), December. COLING.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. cite arxiv:1810.04805Comment: 13 pages.
- Dr Mahmoud El-Haj, et al., editors. (2020). *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, Barcelona, Spain (Online), December. COLING.
- Mahmoud El-Haj, et al., editors. (2021). *Proceedings of the 3rd Financial Narrative Processing Workshop*, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.
- Mahmoud El-Haj, et al., editors. (2022). *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- Gokhan, T., Smith, P., and Lee, M. (2021). Extractive financial narrative summarisation using Sentence-BERT based clustering. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 94–98, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.
- Krimberg, S., Vanetik, N., and Litvak, M. (2021). Summarization of financial documents with TF-IDF weighting of multi-word terms. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 75–80, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July. Association for Computational Linguistics.
- Li, L., Jiang, Y., and Liu, Y. (2020). Extractive financial narrative summarisation based on DPPs. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 100–104, Barcelona, Spain (Online), December. COLING.
- Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July. Association for Computational Linguistics.
- Litvak, M. and Vanetik, N. (2021). Summarization of financial reports with AMUSE. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 31–36, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.
- Litvak, M., Vanetik, N., and Puchinsky, Z. (2020). Hierarchical summarization of financial reports with RUNNER. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 213–225, Barcelona, Spain (Online), December. COLING.
- Mihalcea, R. and Tarau, P. (2004). TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain, July. Association for Computational Linguistics.
- Miller, D. (2019). Leveraging bert for extractive text summarization on lectures.
- Orzhenovskii, M. (2021). T5-LONG-EXTRACT at FNS-2021 shared task. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 67–69, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.
- Zheng, S., Lu, A., and Cardie, C. (2020). SUMSUM@FNS-2020 shared task. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 148–152, Barcelona, Spain (Online), December. COLING.
- Zmandar, N., El-Haj, M., Rayson, P., Abura’Ed, A., Litvak, M., Giannakopoulos, G., Pittaras, N., Carbajo-Coronado, B., and Moreno-Sandoval, A. (2022). The financial narrative summarisation shared task (fns 2022). In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24June. The 13th Language Resources and Evaluation Conference, LREC 2022.