



Universidad Autónoma
de Madrid

Biblos-e Archivo
Repositorio Institucional UAM

Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del capítulo publicado en:
This is an **author produced version** of a book chapter in:

Zavitsanos, E., et al., "The Financial Narrative Summarisation Shared Task (FNS 2023). 2023 IEEE International Conference on Big Data (BigData). Sorrento, Italy: IEEE, 2023. 2890 – 2896

DOI: <https://doi.ieeecomputersociety.org/10.1109/BigData59044.2023.10386228>

Copyright: © 2023 IEEE

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

The Financial Narrative Summarisation Shared Task (FNS 2023)

Elias Zavitsanos

Inst. of Informatics & Telecomm.

NCSR “Demokritos”

P.O. Box 60037, Greece

izavits@iit.demokritos.gr

Aris Kosmopoulos

SciFY PNPC

Athens

Greece

akosmo@scify.org

George Giannakopoulos

Inst. of Informatics & Telecomm.

NCSR “Demokritos”

P.O. Box 60037, Greece

ggianna@iit.demokritos.gr

Marina Litvak

Shamoon College of Engineering

SCE

Beer Sheva, Israel

litvak.marina@gmail.com

Blanca Carbajo-Coronado

Autonomous University of Madrid

LLI-UAM

Madrid, Spain

blanca.carbajo@uam.es

Antonio Moreno-Sandoval

Autonomous University of Madrid

LLI-UAM

Madrid, Spain

antonio.msandoval@uam.es

Mo El-Haj

Lancaster University

SCC

Lancaster, UK

m.el-haj@lancaster.ac.uk

Abstract—This paper presents the results and findings of the Financial Narrative Summarisation Shared Task on summarising UK, Greek, and Spanish annual reports. The shared task was organised as part of the 5th Financial Narrative Processing Workshop (FNP 2023). The Financial Narrative Summarisation Shared Task (FNS 2023) has been running since 2020 as part of the Financial Narrative Processing (FNP) workshop series [15]–[20]. The shared task included one main challenge, which is the use of either abstractive or extractive automatic summarisers to summarise long documents in terms of UK, Greek, and Spanish financial annual reports. This shared task is the fourth to target financial documents. The data for the shared task was created and collected from publicly available annual reports published by firms listed on the Stock Exchanges of the UK, Greece, and Spain. A total number of 6 systems from 3 different teams participated in the shared task.

Index Terms—financial narrative summarisation, annual reports, extractive summarisation, abstractive summarisation, financial narrative processing

I. INTRODUCTION

The task of automatically summarising financial reports presents numerous challenges. Notably, these reports frequently feature an abundance of technical jargon and numerical data, posing formidable obstacles for automated summarisation techniques to grasp their content. Furthermore, financial statements tend to be extensive and intricate, further complicating the identification of pivotal information by automatic summarisation methods. Despite these formidable challenges, the realm of automatic financial narrative summarisation holds great promise.

Financial report summarisation serves as a vital tool for both companies and investors for several compelling reasons. First and foremost, it furnishes a succinct and readily comprehensible overview of a company’s financial performance, enabling investors and analysts to swiftly discern the salient points within a financial report. Additionally, it empowers

companies to furnish a compact overview of their financial performance, a resource of significant value to shareholders and prospective investors alike. Such succinct and comprehensible insights into a company’s financial standing are of paramount importance to investors, aiding them in identifying trends and making informed decisions regarding their investments. Lastly, the practice of financial report summarisation facilitates the early detection of trends and potential issues, affording the opportunity for timely corrective measures.

The summarisation of financial reports assumes a pivotal role for companies and investors seeking to maintain a firm grip on their financial matters. It dispenses crucial information in a concise and easily digestible format, offering the advantage of early issue identification.

In this paper, we introduce the results and discoveries stemming from the Financial Narrative Summarisation Shared Task, which focused on the summarisation of annual reports from the United Kingdom, Greece, and Spain. This collaborative endeavor was seamlessly integrated into the 5th Financial Narrative Processing Workshop (FNP 2023). It is noteworthy that the Financial Narrative Summarisation Shared Task (FNS 2023) has been an ongoing initiative since 2020, firmly embedded within the framework of the Financial Narrative Processing (FNP) workshop series [15]–[20].

A. What are financial narratives

Financial narratives serve as a crucial medium for conveying various aspects of a company’s trajectory, spanning historical, present, and future performance indicators. These narratives also play a pivotal role in elucidating the organisation’s innovative pursuits, research and development strategies, and forthcoming investment initiatives. It is imperative to recognise that the composition and tone of financial narratives can be influenced by a confluence of factors, including the country where the company is listed, the linguistic nuances of re-

porting, and the diverse backgrounds of the board members. However, it is paramount to acknowledge that these narratives are inherently non-objective, primarily owing to managerial reluctance in divulging adverse information about the company. The overarching objective, in most cases, is the enhancement of stock prices and the attraction of requisite funds to propel the realisation of the firm's strategic goals.

B. Categorising Financial Narratives

In the realm of financial narrative analysis, a notable facet emerges in the form of distinct categories of financial communication. These categories encompass Periodic information, Ongoing information, Regulatory information, and Event-related information. Importantly, this classification system serves as a valuable tool for comprehending and interpreting financial narratives across various jurisdictions, such as the United Kingdom, Greece, and Spain.

It is noteworthy that despite the structural diversity in the presentation of financial data in annual reports, a common thread can be discerned across these countries. The PDF reports are long and may lack standardisation, yet companies adhere to a set of guidelines and descriptions that prescribe both the content and the manner of presentation. This adherence to guidelines aids in ensuring a degree of uniformity and coherence in financial narratives, enabling stakeholders to navigate and extract valuable insights from these documents effectively [57].

C. The Imperative of Financial Narrative Summarisation

The motivation behind the summarisation of financial narratives is underscored by their escalating prominence within the life cycle of financial markets. In contemporary times, investors have adopted a practice of scrutinising news articles and engaging in meticulous due diligence processes when contemplating stock acquisitions. Additionally, they rely on a comprehensive analysis of a firm's annual report to assess its potential for growth and anticipate future cash flows. The exponential growth in the volume of available narratives is intrinsically linked to the global proliferation of listed firms, necessitating a sophisticated approach to narrative summarisation and analysis. This growth phenomenon is propelled by the continuous expansion in the number of companies seeking listing on a worldwide scale [58], [59].

II. RELATED WORK

The increased availability of financial reports has been met with research interest in applying automatic summarisation methods. The task of automatic text summarisation aims to produce a condensed, informative, and non-redundant summary from a single or multiple input texts [41]. This is achieved by either identifying and ranking subsets of the input text (i.e., extractive approaches [29] or by generating the summary from scratch (i.e., abstractive methods [38], [52]). Besides this approach-based categorisation, summarisation tasks can also be divided based on the language involved, resulting in single-language, multi-language, and cross-lingual

summarisation. In single-language summarisation input documents and output summaries are in the same language. Multi-language summarisation generates summaries for documents in various languages, with the output summary being in the same language as the input document [54]. Finally, cross-lingual summarisation involves generating a summary in a specific target language when the input document is written in a different language [55], [56].

Extractive methods have been popular for summarising text due to their relative simplicity and the comparatively high requirements of abstractive methods for computational resources and available data. Extractive summarisation utilises scoring approaches to identify and reorder parts of the input (e.g., sentences, phrases, and/or passages), using a variety of feature extraction and evaluation methods [5], [9], [27], [36], [37], [40]. Where adequate data is available, machine learning methods have been employed, such as Hidden Markov Models [25], topic-based modelling [2], genetic algorithms [33], and clustering methods [31], [35], [45].

The employment of summarisation and natural language processing techniques generally has promising applications in the financial domain [13]. The SummariserPort system [8] has been used to produce summaries for financial news, where it utilised lexical cohesion [23], using sentence linkage heuristics to generate the output summary. A summarisation system for financial news was proposed in [22], generating query-based company-tailored summaries. This was done by using unsupervised sentence ranking with simple frequency-based features. Recently, statistical features with heuristic approaches have been used to summarise financial textual disclosures [7], generating summaries with a reduced positive bias, leading to more conservative valuation judgements by investors that receive them. Further, the Financial Narrative Summarisation task [12] of the Multiling 2019 workshop [28] involved the generation of structured summaries from financial narrative disclosures. Considering this body of work, the Financial Narrative Summarisation task (FNS 2020 [15]) resulted in the first large-scale experimental results and state-of-the-art summarisation methods applied to financial data. The task focused on annual reports produced by UK firms listed on the London Stock Exchange (LSE). The shared task was held as part of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (FNP-FNS 2020) [18]. The participating systems used a variety of techniques and methods ranging from rule-based extraction methods [3], [4], [34], [50] to traditional machine learning methods [3], [48], [50] and high-performing deep learning models [1], [3], [4], [32], [47], [50], [51].

One of the main challenges and limitations reported by the participants was the average length of annual reports (around 60,000 words), which made the training process difficult as it requires powerful resources (e.g., GPUs) to avoid long training time. In addition, participants argued that extracting both text and structure from PDF files with numerous tables, charts, and numerical data resulted in noisy data being extracted. Such feedback highlights interesting aspects and challenging com-

ponents of Financial Narrative Summarisation, which presents a high-difficulty task and an interesting research problem that is worth investigating. The 2023 Financial Narrative summarisation shared-task (FNS 2023) promotes this effort by providing such a shared task in the FNP 2023 workshop¹.

III. DATA DESCRIPTION

The Financial Narrative Summarisation (FNS 2023) aims to demonstrate the value and challenges of applying automatic text summarisation to financial text (financial narrative disclosures) written in English, Spanish, and Greek. The task dataset has been extracted from annual financial reports in PDF file format. For the dataset compilation, two to three people had to work for each language.

Figure 1 illustrates an overall summary of the data. For the English data, the average number of words per annual report is around 54K, with half of the reports containing 28K to 70K words. For the Greek data, the average number of words per annual report is around 32,5K, with half of the reports containing 20K to 37K words. For the Spanish data, the average number of words per annual report is around 36,6K, with half of the reports containing 18K to 41K words.

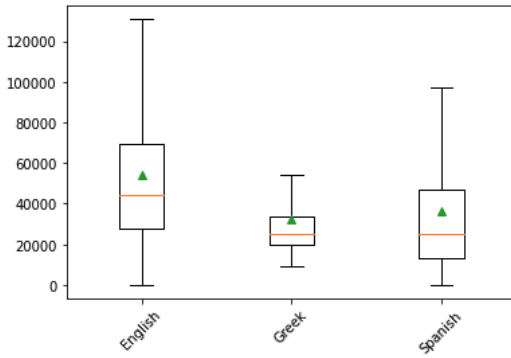


Fig. 1. Dataset statistics per language: word count per text; green triangle: mean value; red horizontal line: median value.

A. English dataset

In the Financial Narrative Summarisation task, we focus on annual reports produced by UK firms listed on The London Stock Exchange (LSE).

In the UK and elsewhere, the structure of the annual report is much less rigid than those produced in the US. Companies produce glossy brochures with a much looser structure, which makes automatic summarisation of narratives in UK annual reports a challenging task.

For the FNS 2023 Shared task² we use approximately 4,000 UK annual reports for firms listed on LSE, covering the period between 2002 and 2022 [10], [12], [20].

We divided the full text within annual reports into **training**, **testing**, and **validation** sets, providing both the full text of each annual report along with gold-standard summaries.

¹Main workshop: <https://wp.lancs.ac.uk/cfie/fnp2023/>

²<http://wp.lancs.ac.uk/cfie/fns2023/>

In total, there are 4,013 annual reports divided into training, testing, and validation sets. Table I shows the dataset details.

TABLE I
FNS 2023 SHARED TASK ENGLISH DATASET

Data type	Train	Validate	Test
Report full text	3,050	413	550
Gold summaries	10,007	1,383	1,804

B. Greek dataset

The Greek dataset is composed of the annual reports for the years 2019 to 2022. These reports are in PDF format and can be 100 to 300 pages long. The Greek reports can be less structured compared to the English ones.

Although the reports seem to follow some pattern, we can observe on several occasions that the structure can differ greatly. For example the “highlights” section can be found in most of the reports but it is not always located in the same sections. Furthermore, some of the reports were problematic during the dataset creation process, and for that reason, they were not used. Common problems were the language used (some were in English), the specific variation of PDF format used, or the very weird structure used by the authors of the report. The initial documents were more than 350, while the final dataset was composed of 312 documents.

The full text was also divided into training, testing, and validation sets in a similar way as with the other datasets. Table II provides the dataset details. The golden summaries were extracted from the statement of the “chairman/board” and the annual report of “management board”.

TABLE II
FNS 2023 SHARED TASK GREEK DATASET

Data type	Train	Validate	Test
Report full text	212	50	50
Gold summaries	424	100	100

C. Spanish dataset

The Spanish dataset is an enhanced version of the FinT-esp corpus [39] and consists of 262 documents with a distribution utterly similar to the Greek dataset (see Table III).

The dates of the annual reports range from 2014 to 2022. The source is in PDF format, with a total number of pages between 40 and 400. In plain text, the files have an average of 36,285 words.

TABLE III
FNS 2023 SHARED TASK SPANISH DATASET

Data type	Train	Validate	Test
Report full text	162	50	50
Gold summaries	324	100	100

The originals were carefully edited by hand, and fragments not containing the narrative (tables, footnotes, headers, etc.) were removed. In addition, the letters from the chairpersons were removed from the reports, as they have been used to make the summaries. Several linguists edited each letter to simplify and reduce the length of the Gold Summaries to 1000 word tokens.

IV. DATA AVAILABILITY

For the shared task we first provide the training and validation sets, which include the full text of each annual report along with the gold-standard summaries. On average, there are at least three gold standard summaries for each annual report with some reports containing up to seven gold standard summaries. The full test set is available only to organisers who evaluate the participating systems. The gold-standard summaries for the test set were not provided to participants in advance.

V. TASK DESCRIPTION

For the purpose of this task, each team was asked to produce one summary for each annual report. The summary length should not exceed 1000 words. We advised that the summary is generated/extracted based on the narrative sections.

Only one summary was allowed for each report, but participating teams were welcome to participate with more than one system. The participants were asked to follow a standard file naming process to aid the automatic evaluation process. Also, for standardisation and consistency, all output summary files were required to be in UTF-8 file format.

Regarding generated outputs from a participant system, the following criteria were requested for each language:

- Each team should produce a no more than 1000 words summary for each annual report in the test set.
- One summary should be provided for each report.
- Each summary should be named following the pattern **ID_summary**. Example: 25082_summary.
- All outputs should be in UTF-8 file format.
- All output summaries should be compressed following the pattern [Team-Name]_[System-name].zip.

A. Evaluation

To evaluate the generated system summaries against the human gold-standard summaries, we used the Java Rouge (JRouge)³ package for ROUGE, using multiple variants (i.e. ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-SU4) [26].

ROUGE is a set of metrics for evaluating automatic text summarisation and machine translation. It works by comparing the produced summary against a set of reference summaries, focusing on the overlapping words (or n-grams) between them. To do this, it uses *Recall*, which indicates how much of the reference summary the system summary captures, and *Precision*, which measures what part of the system summary was in fact relevant.

³<https://github.com/kavgan/ROUGE-2.0>

ROUGE variants include: ROUGE-1, which refers to the overlap of unigrams between the system summary and the reference summary; ROUGE-2 which examines the overlap of bigrams; ROUGE-L which examines the longest in-sequence common n-grams between the system summary and the reference summary, thus not requiring a predefined n-gram length; ROUGE-S considers any pair of words in a sentence in order, allowing for arbitrary gaps (also called “skip-gram co-occurrence”). Finally, ROUGE-SU provides skip-bigram plus unigram-based co-occurrence statistics.

The team with the best ROUGE-2 scores for all three languages was selected as the winner of the competition, based on the assumption that ROUGE-2 has been previously shown to correlate well to human judgement in several settings [60]⁴. The scores are weighted as follows: English (50%), Spanish (25%) and Greek (25%).

VI. DATA SAMPLE

Figure 2 shows the structure of the Financial Narrative Summarisation dataset for all three languages: English, Greek, and Spanish. At the beginning of the shared task, we provided the participants with two directories, corresponding to *training* and *validation* sets. Each contained the full text of the annual reports and the gold standard summaries.

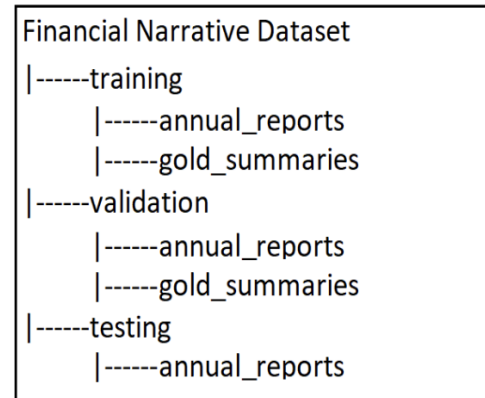


Fig. 2. Dataset structure

The data was provided in plain text format in a directory structure as in Figure 2. Each annual report has a unique ID and it is used across in order to link the full text from an annual report to its gold-standard summaries.

For example, the gold standard summaries for the file called **19** in the training/annual_reports directory can be located in the training_gold_summaries as files with the same ID (19) as a prefix: **19_1** to **19_3**.

VII. BASELINE SUMMARISER

We compared the results of the participating systems to a baseline summariser. We used a BERT extractive summariser

⁴We note that human evaluation in the FNS setting is far from trivial, and remains a challenging research endeavour due to complexity and time requirements.

based on [53]. In particular, we used the multilingual BERT as the starting checkpoint. To handle long texts, we iteratively break the input text into smaller chunks of 512 tokens that are fed to the summariser. Each chunk’s summaries are concatenated to produce the final summary of the text. The process continues until the final summary of the text has a size less or equal to 1000 tokens. Figure 3 illustrates this process.

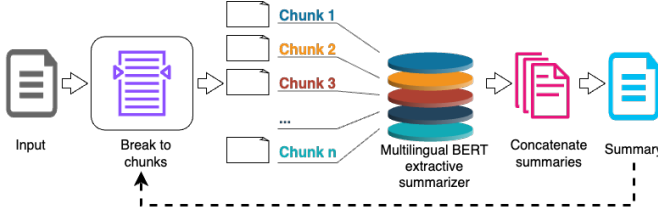


Fig. 3. The baseline summarisation system.

VIII. PARTICIPANTS AND SYSTEMS

In total, 6 summarisation systems by 3 different teams have participated and submitted their system summaries to FNS 2023. The teams are presented in Table IV.

TABLE IV
FNS 2023 PARTICIPATING TEAMS AND THEIR AFFILIATIONS

Team	Affiliation
Rocky	MBZUAI
SCE	Sami Shamoon College of Engineering
SSC_AI_RG	State Street Corporation

The **SCE** team employed the Positional Language Model (PLM) for all three datasets. The team submitted two systems—one performs a pre-selection of the first 10% of a document before the PLM application, and another also employs the BERTSUM summariser that selects text with a length of 3000 words. Both options are motivated by reducing the space and time complexity of PLM and by removing redundant and noisy content from a long report. Since PLM is a statistical approach, it does not require adjustment for different languages. Therefore, the same systems (including those that employ BERTSUM extractive summariser) were applied to all three languages.

The **Rocky** team employed the T5 encoder-decoder pre-trained model for the English dataset, while for the Greek and Spanish, they utilised a fine-tuned MT5 model, a multilingual T5 model pre-trained on a corpus that includes 101 languages.

The **SSC_AI_RG** team implemented an algorithm for allocating K words in narrative sections or areas using a weighted-based approach. In the context of extractive summarisation, they experimented with Top-K, BERT, and Bart extractive summarisers. In order to identify key narrative sections, a classification system decides if a section should be in the summary or not. Clusters were created around narrative sentences based on language independence, structure independence, and neighborhood assumptions.

IX. RESULTS AND DISCUSSION

The participating systems used a variety of techniques and methods ranging from fine-tuning pre-trained transformers to using high-performing deep learning models and word embeddings.

In addition, the participating teams used methods to investigate the hierarchy of the annual reports to try and detect structure and extract the narrative sections, in order to identify the parts in the report from which the gold summaries were extracted.

The majority of the applied techniques were extractive since the dataset is highly structured with discrete sections.

The results in Table V show the ROUGE-2 F measure score for each language. The systems are ranked according to the final score which is weighted as follows: English (50%), Spanish (25%), and Greek (25%). Please note that we use 0.00 to indicate a no-participation for a given language.

TABLE V
FNS 2023 RESULTS. EN: ENGLISH, ES: SPANISH, EL: GREEK

Team_System	EN	EL	ES	Score
SSC_AI_RG_DiMSum	0,32	0,12	0,11	0,217
SCE_plm	0,27	0,13	0,14	0,202
SCE_bertplm	0,26	0,13	0,13	0,195
SSC_AI_RG_GenAI_DiMSum2	0,24	0,00	0,00	0,120
Rocky_T5	0,12	0,13	0,08	0,112
SSC_AI_RG_GenAI_DiMSum1	0,22	0,00	0,00	0,110
MBertExtractive_baseline	0,11	0,05	0,08	0,087

The results show a very close performance of the first three systems. Specifically, the team SSC_AI_RG with the *DimSum* system ranked first, although very close to the systems of team SCE. Another observation is that the baseline system that was developed as a comparison against the participants ranks last, although two of the systems did not participate in the Greek and Spanish datasets.

The complete evaluation results, including ROUGE 1, 2, L, and SU4, are shown in tables VI, VII, and VIII. Such results will be used as a comparison line in the future, by incorporating them into a venue of results, techniques, and approaches, which we hope will be helpful to researchers working on Financial Text Summarisation.

REFERENCES

- [1] Agarwal, R., Verma, I., and Chatterjee, N. (2020). Lang research lab at fincausal 2020, task 1: A knowledge induced neural net for causality detection. In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pages 33–39.
- [2] Aries, A., Zegour, D. E., and Hidouci, K. W. (2015). All summariser system at multiling 2015: Multilingual single and multi-document summarisation. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 237–244. The Association for Computer Linguistics.
- [3] Arora, P. and Radhakrishnan, P. (2020). Amex ai-labs: An investigative study on extractive summarisation of financial documents. In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pages 137–142.

TABLE VI
FNS 2023 RESULTS. ALL ROUGE SCORES IN THE ENGLISH DATASET.

Team_System	R-1/P	R-1/R	R-1/F	R-2/P	R-2/R	R-2/F	R-L/P	R-L/R	R-L/F	R-SU4/P	R-SU4/R	R-SU4/F
SCE_bertplm	0,43	0,44	0,42	0,26	0,28	0,26	0,38	0,45	0,41	0,30	0,33	0,30
SCE_plm	0,44	0,46	0,43	0,28	0,30	0,27	0,39	0,46	0,41	0,33	0,35	0,33
Rocky_T5	0,27	0,25	0,24	0,11	0,15	0,12	0,50	0,19	0,26	0,12	0,19	0,14
SSC_AI_RG_DiMSum	0,48	0,53	0,48	0,27	0,44	0,32	0,46	0,51	0,47	0,24	0,48	0,31
SSC_AI_RG_DiMSum1	0,48	0,37	0,38	0,23	0,24	0,22	0,44	0,36	0,37	0,23	0,29	0,24
SSC_AI_RG_DiMSum2	0,38	0,48	0,40	0,21	0,33	0,24	0,36	0,44	0,38	0,24	0,38	0,28
MBertExtractive_baseline	0,31	0,27	0,28	0,11	0,14	0,11	0,23	0,24	0,23	0,12	0,19	0,14

TABLE VII
FNS 2023 RESULTS. ALL ROUGE SCORES IN THE GREEK DATASET.

Team_System	R-1/P	R-1/R	R-1/F	R-2/P	R-2/R	R-2/F	R-L/P	R-L/R	R-L/F	R-SU4/P	R-SU4/R	R-SU4/F
SCE_bertplm	0,44	0,25	0,32	0,21	0,10	0,13	0,36	0,21	0,26	0,25	0,13	0,17
SCE_plm	0,44	0,25	0,32	0,21	0,10	0,13	0,36	0,21	0,26	0,27	0,13	0,17
Rocky_T5	0,48	0,23	0,31	0,19	0,10	0,13	0,35	0,20	0,25	0,23	0,12	0,16
SSC_AI_RG_DiMSum	0,44	0,22	0,29	0,25	0,08	0,12	0,33	0,16	0,20	0,26	0,12	0,16
MBertExtractive_baseline	0,26	0,19	0,22	0,08	0,04	0,05	0,10	0,11	0,10	0,12	0,09	0,10

TABLE VIII
FNS 2023 RESULTS. ALL ROUGE SCORES IN THE SPANISH DATASET.

Team_System	R-1/P	R-1/R	R-1/F	R-2/P	R-2/R	R-2/F	R-L/P	R-L/R	R-L/F	R-SU4/P	R-SU4/R	R-SU4/F
SCE_bertplm	0,34	0,51	0,36	0,12	0,20	0,13	0,20	0,27	0,21	0,17	0,27	0,18
SCE_plm	0,40	0,46	0,41	0,14	0,15	0,14	0,25	0,27	0,25	0,20	0,21	0,20
Rocky_T5	0,45	0,36	0,39	0,09	0,08	0,08	0,15	0,16	0,15	0,16	0,14	0,14
SSC_AI_RG_DiMSum	0,37	0,45	0,40	0,11	0,13	0,11	0,16	0,17	0,16	0,16	0,20	0,17
MBertExtractive_baseline	0,38	0,36	0,36	0,08	0,08	0,08	0,14	0,14	0,14	0,15	0,15	0,15

- [4] Azzi, A. A. and Kang, J. (2020). Extractive summarisation system for annual reports. In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pages 143–147.
- [5] Baxendale, P. B. (1958). Machine-made index for technical literature—an experiment. IBM Journal of research and development, 2(4):354–361.
- [6] Beltagy, I., Peters, M. E., and Cohan, A. (2020). Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
- [7] Cardinaels, E., Hollander, S., and White, B. J. (2019). Automatic summarisation of earnings releases: attributes and effects on investors’ judgments. Review of Accounting Studies, 24(3):860–890.
- [8] de Oliveira, P. C. F., Ahmad, K., and Gillam, L. (2002). A financial news summarisation system based on lexical cohesion. In Proceedings of the International Conference on Terminology and Knowledge Engineering, Nancy, France.
- [9] Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM (JACM), 16(2):264–285.
- [10] El-Haj, M., Rayson, P., Young, S., and Walker, M. (2014). Detecting document structure in a very large corpus of UK financial reports. In European Language Resources Association (ELRA).
- [11] El-Haj, M., Rayson, P., and Moore, A. (2018). The first financial narrative processing workshop (fnp 2018). In Proceedings of the LREC 2018 Workshop.
- [12] El-Haj, M., Rayson, P., Alves, P., Herrero-Zorita, C., and Young, S. (2019a). Multilingual financial narrative processing: Analysing annual reports in English, spanish and portuguese. World Scientific Publishing.
- [13] El-Haj, M., Rayson, P., Walker, M., Young, S., and Simaki, V. (2019b). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. Journal of Business Finance & Accounting, 46(3-4):265–306.
- [14] Mahmoud El-Haj, et al., editors. (2019c). Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019), Turku, Finland, September. Linköping University Electronic Press.
- [15] El-Haj, M., AbuRa’ed, A., Litvak, M., Pittaras, N., and Giannakopoulos, G. (2020a). The financial narrative summarisation shared task (FNS 2020). In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, pages 1–12, Barcelona, Spain (Online), December. COLING.
- [16] El-Haj, M., Zmandar, N., Rayson, P., AbuRa’ed, A., Litvak, M., Pittaras, N., Giannakopoulos, G., Kosmopoulos, A., Carbajo-Coronado, B., Moreno-Sandoval, A. The Financial Narrative Summarisation Shared Task (FNS 2022). Proceedings of the 4th Financial Narrative Processing Workshop, LREC 2022, pages 43–52.
- [17] Mahmoud El-Haj, et al., editors. (2020b). Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, Barcelona, Spain (Online), December. COLING.
- [18] El-Haj, M., Athanasakou, V., Ferradans, S., Salzedo, C., Elhag, A., Bouamor, H., Litvak, M., Rayson, P., Giannakopoulos, G., and Pittaras, N. (2020c). Proceedings of the 1st joint workshop on financial narrative processing and multiling financial summarisation. In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation.
- [19] 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.
- [20] 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.
- [21] El-Haj, M. (2019). Multiling 2019: Financial narrative summarisation. In Proceedings of the Workshop MultiLing 2019: Summarisation Across Languages, Genres and Sources, pages 6–10.
- [22] Filippova, K., Surdeanu, M., Ciaramita, M., and Zaragoza, H. (2009). Company-oriented extractive summarisation of financial news. In Pro-

- ceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 246–254.
- [23] Flowerdew, J. and Mahlberg, M. (2009). *Lexical cohesion and corpus linguistics*, volume 17. John Benjamins Publishing.
- [24] Foroutan, N., Romanou, A., Massonnet, S., Lebre, R., and Aberer, K. (2022). Multilingual text summarisation on financial documents. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [25] Fung, P. and Ngai, G. (2006). One story, one flow: Hidden markov story models for multilingual multidocument summarisation. *ACM Transactions on Speech and Language Processing (TSLP)*, 3(2):1–16.
- [26] Ganesan, K. (2018). Rouge 2.0: Updated and improved measures for evaluation of summarisation tasks. *arXiv preprint arXiv:1803.01937*.
- [27] Giannakopoulos, G., Karkaletsis, V., Vouros, G., and Stamatoopoulos, P. (2008). Summarisation system evaluation revisited: N-gram graphs. *ACM Transactions on Speech and Language Processing (TSLP)*, 5(3):1–39.
- [28] Giannakopoulos, G. (2019). Proceedings of the workshop multiling 2019: Summarisation across languages, genres and sources. In *Proceedings of the Workshop MultiLing 2019: Summarisation Across Languages, Genres and Sources*.
- [29] Gupta, V. and Lehal, G. S. (2010). A survey of text summarisation extractive techniques. *Journal of emerging technologies in web intelligence*, 2(3):258–268.
- [30] Khanna, U., Ghodrathnama, S., Moll’a, D., and Beheshti, A. (2022). Transformer-based models for long document summarisation in financial domain. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [31] Kruengkrai, C. and Jaruskulchai, C. (2003). Generic text summarisation using local and global properties of sentences. In *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, pages 201–206. IEEE.
- [32] La Quatra, M. and Cagliero, L. (2020). End-to-end training for financial report summarisation. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 118–123.
- [33] Litvak, M., Last, M., and Friedman, M. (2010). A new approach to improving multilingual summarisation using a genetic algorithm. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 927–936.
- [34] Litvak, M., Vanetik, N., and Puchinsky, Z. (2020). Sce-summary at the fns 2020 shared task. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 124–129.
- [35] Liu, S. and Lindroos, J. (2006). Experiences from automatic summarisation of imf staff reports. *Practical Data Mining: Applications, Experiences and Challenges*, page 43.
- [36] Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.
- [37] McCargar, V. (2004). Statistical approaches to automatic text summarisation. *Bulletin of the American Society for Information Science and Technology*, 30(4):21–25.
- [38] Moratanch, N. and Chitrakala, S. (2016). A survey on abstractive text summarisation. In *2016 International Conference on Circuit, power and computing technologies (ICCPCT)*, pages 1–7. IEEE.
- [39] Moreno-Sandoval, A., Gisbert, A., and Montoro, H. (2020). Fint-esp: a corpus of financial reports in spanish. In Fuster, et al., editors, *Multiperspectives in analysis and corpus design*, pages 89–102, Granada. Comares.
- [40] Mori, T. (2002). Information gain ratio as term weight: the case of summarisation of ir results. In *COLING 2002: The 19th International Conference on Computational Linguistics*.
- [41] Nenkova, A. and McKeown, K. (2011). *Automatic summarisation*. Now Publishers Inc.
- [42] Ogden, A. and El-Haj, M. (2022). Financial narrative summarisation using a hybrid tf-idf and clustering summariser: Ao-lancs system at fns 2022. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [43] 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.
- [44] Pant, M. and Chopra, A. (2022). Multilingual financial documentation summarisation by team tredence for fns2022. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [45] Radev, D., Jing, H., and Budzikowska, M. (2000). Centroid-based summarisation of multiple documents: Clustering, sentence extraction, and evaluation. In *Proceedings of the ANLP/NAACL-2000 Workshop on Summarisation*.
- [46] Shukla, N. K., Vaid, A., Katikeri, R. C., Keeriyadath, S., and Raja, M. (2022). Dimsum: Distributed and multilingual summarisation of financial narratives. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [47] Singh, A. (2020). Point-5: Pointer network and t-5 based financial narrative summarisation. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 105–111.
- [48] Suarez, J. B., Martínez, P., and Martínez, J. L. (2020). Combining financial word embeddings and knowledge-based features for financial text summarisation uc3m-mc system at fns-2020. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 112–117.
- [49] Vaca, A., Segurado, A., Betancur, D., and Jiménez, A. B. (2022). Extractive and abstractive summarisation methods for financial narrative summarisation in english, spanish and greek. In *Proceedings of the 4th Financial Narrative Processing Workshop*, Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.
- [50] Vhatkar, A., Bhattacharyya, P., and Arya, K. (2020). Knowledge graph and deep neural network for extractive text summarisation by utilising triples. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 130–136.
- [51] Zheng, S., Lu, A., and Cardie, C. (2020). Sum-sum at fns-2020 shared task. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 148–152.
- [52] Zmandar, N., Singh, A., El-Haj, M., and Rayson, P. (2021). Joint abstractive and extractive method for long financial document summarisation. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 99–105, Lancaster, United Kingdom, 15–16 September. Association for Computational Linguistics.
- [53] Derek Miller. 2019. Leveraging BERT for extractive text summarisation on lectures. *arXiv preprint arXiv:1906.04165* (2019).
- [54] Alkesh Patel, Tanveer Siddiqui, and US Tiwary. A language independent approach to multilingual text summarisation. *Large scale semantic access to content (text, image, video, and sound)*, pages 123–132, 2007.
- [55] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 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 2020.
- [56] Constantin Orăsan and Oana Andreea Chiorean. Evaluation of a cross-lingual Romanian-English multi-document summariser. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA).
- [57] El-Haj, M., Alves, P., Rayson, P., Walker, M. and Young, S., 2020. Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as PDF files. *Accounting and Business Research*, 50(1), pp.6-34.
- [58] El-Haj, M., Rayson, P., El Maarouf, I., Bentabet, N.I., Mariko, D., Labidurie, E., Litvak, M., Giannakopoulos, G., AbuRa’ed, A. and Zmandar, N., 2021. Review of the State of the Art in Financial Narrative Processing.
- [59] El-Haj, M., Sandoval, A.M. and Millán, J.A.J., 2021. Machine learning models for classifying Spanish beaters and non-beaters financial reports. In *Financial narrative processing in spanish* (pp. 179-198). Tirant Humanidades.
- [60] Graham, Yvette, 2015. Re-evaluating automatic summarization with BLEU and 192 shades of ROUGE. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 128-137).