

The Financial Narrative Summarisation Shared Task (FNS 2022)

Mahmoud El-Haj¹, Nadhem Zmandar¹, Paul Rayson¹, Ahmed AbuRa'ed²,
Marina Litvak³, Nikiforos Pittaras⁴, George Giannakopoulos^{4,5}, Aris Kosmopoulos^{4,5},
Blanca Carbajo-Coronado⁶, Antonio Moreno-Sandoval⁶

¹Lancaster University, UK

²University of British Columbia, Canada

³Shamoon College of Engineering, Israel

⁴NCSR “Demokritos”, Athens, Greece

⁵SciFY PNPC, Athens, Greece

⁶Universidad Autónoma Madrid, Spain

¹{m.el-haj, n.zmandar, p.rayson}@lancaster.ac.uk

²ahmed.aburaed@upf.edu, ³marinal@ac.sce.ac.il

^{4,5}ggianna@iit.demokritos.gr, pittarasnikif@gmail.com, akosmo@scify.org

⁶{blanca.carbajo, antonio.msandoval}@uam.es

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 Financial Narrative Processing 2022 Workshop (**FNP 2022 Workshop**). The Financial Narrative Summarisation Shared Task (FNS-2022) has been running since 2020 as part of the Financial Narrative Processing (FNP) workshop series (El-Haj et al., 2022; El-Haj et al., 2021; El-Haj et al., 2020b; El-Haj et al., 2019c; El-Haj et al., 2018). The shared task included one main task 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 third 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 UK, Greece and Spain. A total number of 14 systems from 7 different teams participated in the shared task.

1. What are financial narratives

Companies produce a variety of reports containing both narrative and numerical information at various times during their financial year, including annual financial reports. This creates vast amounts of financial information which can be impossible to navigate, handle and keep track of. This shows the vital need for automatic summarisation systems in order to reduce the time and effort of both the shareholders and investors.

2. Related Work

The increased availability of financial reports data has been met with research interest for applying automatic summarisation methods. The task of automatic text summarisation aims to produce a condensed, informative and non-redundant summaries from a single or multiple input texts (Nenkova and McKeown, 2011). This is achieved by either identifying and ranking subsets of the input text (i.e. extractive approaches ((Gupta and Lehal, 2010)), or by generating the summary from scratch (i.e. abstractive methods (Moratanch and Chitrakala, 2016; Zmandar et al., 2021)). Extractive methods have been a popular venue 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 fea-

ture extraction and evaluation methods (Luhn, 1958; Baxendale, 1958; Edmundson, 1969; Mori, 2002; McCargar, 2004; Giannakopoulos et al., 2008). Where adequate data is available, machine learning methods have been employed, such as Hidden Markov Models (Fung and Ngai, 2006), topic-based modelling (Aries et al., 2015), genetic algorithms (Litvak et al., 2010) and clustering methods (Radev et al., 2000; Liu and Lindroos, 2006; Kruengkrai and Jaruskulchai, 2003).

The employment of summarisation and natural language processing techniques in general has promising applications in the financial domain (El-Haj et al., 2019b). The SummariserPort system (de Oliveira et al., 2002) has been used to produce summaries for financial news, where it utilized lexical cohesion (Flowerdew and Mahlberg, 2009), using sentence linkage heuristics to generate the output summary. A summarisation system for financial news was proposed in (Filippova et al., 2009) generating query-based company-tailored summaries. This was done through using unsupervised sentence ranking with simple frequency-based features. Recently, statistical features with heuristic approaches have been used to summarise financial textual disclosures (Cardinaels et al., 2019), generating summaries with reduced positive bias, leading to more conservative valuation judgements by investors that receive them. Further, 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. Considering this body of work, the Financial Narrative Summarisation task (FNS 2020 (El-Haj et al., 2020a)) task 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) (El-Haj et al., 2020c). The participating systems used a variety of techniques and methods ranging from rule based extraction methods (Litvak et al., 2020; Vhatkar et al., 2020; Arora and Radhakrishnan, 2020; Azzi and Kang, 2020) to traditional machine learning methods (Suarez et al., 2020; Vhatkar et al., 2020; Arora and Radhakrishnan, 2020) and high performing deep learning models (Agarwal et al., 2020; Singh, 2020; La Quatra and Cagliero, 2020; Vhatkar et al., 2020; Arora and Radhakrishnan, 2020; Azzi and Kang, 2020; Zheng et al., 2020).

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 components of Financial Narrative Summarisation, which presents a high-difficulty task and an interesting research problem that is worth investigating. The 2022 Financial Narrative summarisation task (FNS 2022) promotes this effort by providing such a shared task in the FNP 2022 workshop¹.

3. Data Description

The Financial Narrative Summarisation (FNS 2022) aims to demonstrate the value and challenges of applying automatic text summarisation to financial text written in English, Spanish and Greek, usually referred to as financial narrative disclosures. The task dataset has been extracted from UK, Greek and Spanish annual reports published in PDF file format.

3.1. 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, annual report structure 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.

¹Main workshop: <http://wp.lancs.ac.uk/cfie/fnp2022/>

For the FNS 2022 Shared task² we use approximately 4,000 UK annual reports for firms listed on LSE, covering the period between 2002 and 2017 (El-Haj et al., 2014; El-Haj et al., 2019a).

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.

In total there are 3,863 annual reports divided into training, testing and validation sets. Table 1 shows the dataset details.

Data Type	Train	Validate	Test
Report full text	3,000	363	500
Gold summaries	9,873	1,250	1,673

Table 1: FNS 2022 Shared Task Dataset

3.2. Greek Dataset

The Greek dataset is composed by the annual reports of years 2019 and 2020. These reports are in PDF format and can be from 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 at 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 at the same sections. Furthermore some of the reports were problematic during the dataset creation process and 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 around 300, while the final dataset was composed by 262 documents.

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

Table 2: FNS 2022 Shared Task Greek Dataset

The full text was also divided into training, testing and validation sets in a similar way as with the other datasets. Table 2 shows the dataset details. The golden summaries were extracted from the statement of the “chairman/board” and the annual report of “management board”.

3.3. Spanish Dataset

The Spanish dataset is taken from the FinT-esp corpus (Moreno-Sandoval et al., 2020) and consists of 262 documents with a distribution utterly similar to the Greek dataset (see Table 3).

²<http://wp.lancs.ac.uk/cfie/fns2022/>

The dates of the annual reports range from 2014 to 2018. 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.

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

Table 3: FNS 2022 Shared Task Spanish Dataset

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.

4. 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.

5. 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 testing 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>_Summaries.tar.gz`.

5.1. 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) (Ganesan, 2018).

The team with the best ROUGE-2 scores for all three languages was selected as the winner of the competition. The scores are weighted as follows: English (50%), Spanish (25%) and Greek (25%) as later shown in Table 5.

6. Data Sample

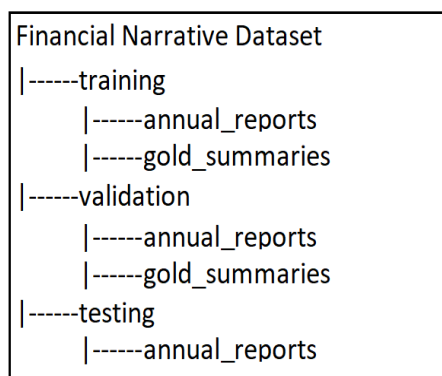


Figure 1: Dataset Structure

Figure 1 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.

The data was provided in plain text format in a directory structure as in Figure 1. 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**.

7. Participants and Systems

In total, 14 summarisation systems by 7 different teams have participated and submitted their system summaries to FNS 2022, the teams are presented in Table 4.

AO-Lancs team produced a hybrid summariser using TF-IDF and clustering methodology. Utilising statistical methods to combine the TF-IDF Sentence score with the Clustering Euclidean distance for each sentence, producing new hybrid sentence rankings. A

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

Team	Affiliation
LSIR	École Polytechnique Fédérale de Lausanne (EPFL)
SSC-AI-RG	State Street Corporation
IIC	Instituto de Ingeniería de Conocimiento
TREDENCE	Tredence Inc.
LIPI	Fidelity Investments, Jadavpur University
MACQUARIE	Macquarie University
AO-LANCS	Lancaster University

Table 4: FNS 2022 participating teams and their affiliations

60/40 weighting in favour of clustering was applied when combining the scores (Ogden and El-Haj, 2022).

LSIR team participated with two systems; the first uses a pre-trained multilingual abstractive summarisation model (mT5) that was fine-tuned on the downstream task to generate the start of the summaries, while the second system approaches the problem as an extractive summariser in which a similarity search is performed on the trained span embeddings to find good candidates for a summary start. The language models were fine-tuned on a financial document collection of three languages; English, Spanish and Greek, and aim to identify the beginning of the summary narrative part of the document. The system based on mT5 achieves the highest performance in the given task, ranked 1st on Rouge scores over the three languages (Foroutan et al., 2022).

Tredence team submitted a multi-lingual long document summarisation system. They developed task-specific summarisation methods for all three languages: English, Spanish and Greek. The solution is divided into two parts, where a RoBERTa model was fine tuned to identify and extract summarising segments from English documents and T5 based models were used for summarising Spanish and Greek documents. An mT5 model was fine-tuned to identify potential narrative sections for Greek and Spanish, followed by fine tuning mT5 and T5 (Spanish version) for abstractive summarisation task. This system also features a novel approach for generating summarisation training dataset using long document segmentation and the semantic similarity across segments (Pant and Chopra, 2022).

SSC_AI_RG team created an algorithm called K-Maximal Word Allocation which allocates K words i.e. 1000 words in narrative sections or areas according to their weights as amount of words to be generated from a section. For extraction we experimented with Top-K, Bert and Bart extractive summarisers. To identify key narrative sections in English reports, they built a section classification system which classifies if the section should be in summary or not. They extracted TOC, section names and applied lookup in summaries to annotate section names. Clusters were

created around narrative sentences based on following assumptions: Language Independence, Structure Independence and Neighbourhood Assumption. Top M Narrative Sections according to their weights were translated to Spanish and Greek. Keywords were extracted from these with weights later to be used to identify narrative sentences and areas and calculate weights (Shukla et al., 2022).

LIPI team has used the system provided by last year’s winning team (Orzhenovskii, 2021), the original summariser provided by Orzhenovskii relies on T5 in order to perform the summarisation.

IIC team developed a summariser based on a sequence classification task whose objective was to find the sentence where the summary begins in the English dataset. For the reports in Spanish and Greek they used an abstractive strategy creating an Encoder-Decoder architecture in Spanish, MariMari, based on an existing Encoding-only model; they also trained multilingual Encoder-Decoder models for this task. As for the Greek dataset, they created a translation-summary-translation system in which the reports were translated into English and summarised, and then the summaries were translated back to Greek (Vaca et al., 2022).

Finally, **Macquarie team** used Longformer-Encoder-Decoder (LED) (Beltagy et al., 2020) model to generate the summaries. They also investigated the multi-stage fine-tuning approach to explore if it helps the model to generate better on the financial domain and avoids the problem of forgetting (Khanna et al., 2022).

8. 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.

Team	En	ES	EL	Score
LSIR-1	0.37	0.16	0.14	0.26
SSC-AI-RG-1	0.33	0.15	0.19	0.25
SSC-AI-RG-3	0.32	0.15	0.19	0.24
IIC	0.37	0.13	0.10	0.24
SSC-AI-RG-2	0.30	0.15	0.19	0.23
TREDENCE-2	0.32	0.13	0.14	0.23
TREDENCE-1	0.32	0.13	0.14	0.23
LIPI	0.38	0.07	0.05	0.22
TREDENCE-3	0.32	0.13	0.07	0.21
LSIR-3	0.28	0.14	0.13	0.21
MACQUARIE-1	0.30	0.00	0.00	0.15
MACQUARIE-3	0.30	0.00	0.00	0.15
MACQUARIE-2	0.30	0.00	0.00	0.15
AO-LANCS	0.14	0.13	0.13	0.14

Table 5: FNS 2022 results
EN: English, ES: Spanish, EL: Greek

The results in Table 5 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%). The results shows that Team LSIR ranked first using the first run of their module. Please note that we use 0.00 to indicate a no-participation for a given language.

The complete evaluation results including ROUGE 1, 2, L and SU4 are show in tables 6, 7, 8, 9,10 and 11. 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 useful to researchers working on Financial Text Summarisation.

9. Bibliographical References

- Agarwal, R., Verma, I., and Chatterjee, N. (2020). Langresearchlab_nc 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.
- Aries, A., Zegour, D. E., and Hidouci, K. W. (2015). Allsummarizer system at multiling 2015: Multilingual single and multi-document summarization. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 237–244. The Association for Computer Linguistics.
- Arora, P. and Radhakrishnan, P. (2020). Amex ai-labs: An investigative study on extractive summarization of financial documents. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 137–142.
- Azzi, A. 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.
- Baxendale, P. B. (1958). Machine-made index for technical literature—an experiment. *IBM Journal of research and development*, 2(4):354–361.
- Beltagy, I., Peters, M. E., and Cohan, A. (2020). Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.
- Cardinaels, E., Hollander, S., and White, B. J. (2019). Automatic summarization of earnings releases: attributes and effects on investors’ judgments. *Review of Accounting Studies*, 24(3):860–890.
- de Oliveira, P. C. F., Ahmad, K., and Gillam, L. (2002). A financial news summarization system based on lexical cohesion. In *Proceedings of the International Conference on Terminology and Knowledge Engineering*, Nancy, France.
- Edmundson, H. P. (1969). New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2):264–285.
- 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)*.
- El-Haj, M., Rayson, P., and Moore, A. (2018). The first financial narrative processing workshop (fnp 2018). In *Proceedings of the LREC 2018 Workshop*.
- 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*.
- 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.
- Mahmoud El-Haj, et al., editors. (2019c). *Proceedings of the Second Financial Narrative Processing Workshop (FNP 2019)*, Turku, Finland, September. Linköping University Electronic Press.
- 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.
- 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.
- 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 Fi-*

- Financial Narrative Processing and MultiLing Financial Summarisation*.
- 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.
- El-Haj, M. (2019). Multiling 2019: Financial narrative summarisation. In *Proceedings of the Workshop MultiLing 2019: Summarization Across Languages, Genres and Sources*, pages 6–10.
- Filippova, K., Surdeanu, M., Ciaramita, M., and Zaragoza, H. (2009). Company-oriented extractive summarization of financial news. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 246–254.
- Flowerdew, J. and Mahlberg, M. (2009). *Lexical cohesion and corpus linguistics*, volume 17. John Benjamins Publishing.
- Foroutan, N., Romanou, A., Massonnet, S., Lebre, R., and Aberer, K. (2022). Multilingual text summarization 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.
- Fung, P. and Ngai, G. (2006). One story, one flow: Hidden markov story models for multilingual multidocument summarization. *ACM Transactions on Speech and Language Processing (TSLP)*, 3(2):1–16.
- Ganesan, K. (2018). Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. *arXiv preprint arXiv:1803.01937*.
- Giannakopoulos, G., Karkaletsis, V., Vouros, G., and Stamatopoulos, P. (2008). Summarization system evaluation revisited: N-gram graphs. *ACM Transactions on Speech and Language Processing (TSLP)*, 5(3):1–39.
- Giannakopoulos, G. (2019). Proceedings of the workshop multiling 2019: Summarization across languages, genres and sources. In *Proceedings of the Workshop MultiLing 2019: Summarization Across Languages, Genres and Sources*.
- Gupta, V. and Lehal, G. S. (2010). A survey of text summarization extractive techniques. *Journal of emerging technologies in web intelligence*, 2(3):258–268.
- 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.
- Kruengkrai, C. and Jaruskulchai, C. (2003). Generic text summarization using local and global properties of sentences. In *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, pages 201–206. IEEE.
- La Quatra, M. and Cagliero, L. (2020). End-to-end training for financial report summarization. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 118–123.
- Litvak, M., Last, M., and Friedman, M. (2010). A new approach to improving multilingual summarization using a genetic algorithm. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 927–936.
- 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.
- Liu, S. and Lindroos, J. (2006). Experiences from automatic summarization of imf staff reports. *Practical Data Mining: Applications, Experiences and Challenges*, page 43.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.
- McCargar, V. (2004). Statistical approaches to automatic text summarization. *Bulletin of the American Society for Information Science and Technology*, 30(4):21–25.
- Moratanch, N. and Chitrakala, S. (2016). A survey on abstractive text summarization. In *2016 International Conference on Circuit, power and computing technologies (ICCPCT)*, pages 1–7. IEEE.
- 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.
- Mori, T. (2002). Information gain ratio as term weight: the case of summarization of ir results. In *COLING 2002: The 19th International Conference on Computational Linguistics*.
- Nenkova, A. and McKeown, K. (2011). *Automatic summarization*. Now Publishers Inc.
- 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.
- 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.
- Pant, M. and Chopra, A. (2022). Multilingual financial

- documentation summarization 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.
- Radev, D., Jing, H., and Budzikowska, M. (2000). Centroid-based summarization of multiple documents: Clustering, sentence extraction, and evaluation. In *Proceedings of the ANLP/NAACL-2000 Workshop on Summarization*.
- Shukla, N. K., Vaid, A., Katikeri, R. C., Keeriyadath, S., and Raja, M. (2022). Dimsum: Distributed and multilingual summarization 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.
- 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.
- Suarez, J. B., Martínez, P., and Martínez, J. L. (2020). Combining financial word embeddings and knowledge-based features for financial text summarization uc3m-mc system at fns-2020. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 112–117.
- Vaca, A., Segurado, A., Betancur, D., and Jiménez, A. B. (2022). Extractive and abstractive summarization methods for financial narrative summarization 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.
- Vhatkar, A., Bhattacharyya, P., and Arya, K. (2020). Knowledge graph and deep neural network for extractive text summarization by utilizing triples. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 130–136.
- 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.
- Zmandar, N., Singh, A., El-Haj, M., and Rayson, P. (2021). Joint abstractive and extractive method for long financial document summarization. In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 99–105, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.

Appendix A - English Task Results

Model	R-1 / R	R-1 / P	R-1 / F	R-2 / R	R-2 / P	R-2 / F
LIPI	0.587	0.451	0.496	0.472	0.326	0.374
IIC	0.566	0.472	0.497	0.438	0.337	0.366
LSIR-1	0.583	0.443	0.489	0.464	0.317	0.365
SSC-AI-RG-1	0.551	0.482	0.495	0.455	0.272	0.327
TREDENCE-2	0.49	0.461	0.462	0.346	0.323	0.322
TREDENCE-3	0.49	0.461	0.462	0.346	0.323	0.322
SSC-AI-RG-3	0.524	0.483	0.484	0.421	0.274	0.319
TREDENCE-1	0.428	0.503	0.447	0.305	0.363	0.317
MACQUARIE-1	0.48	0.438	0.443	0.334	0.302	0.303
MACQUARIE-3	0.48	0.435	0.442	0.333	0.301	0.302
MACQUARIE-2	0.476	0.434	0.441	0.33	0.297	0.301
SSC-AI-RG-2	0.472	0.491	0.462	0.358	0.282	0.3
LSIR-3	0.49	0.442	0.451	0.355	0.241	0.275
AO-LANCS	0.372	0.292	0.317	0.184	0.126	0.143

Table 6: ROUGE-1 and ROUGE-2 on English dataset ordered by R2 F1 score

Model	R-L / R	R-L / P	R-L / F	R-SU4 / R	R-SU4 / P	R-SU4 / F
LIPI	0.559	0.449	0.487	0.515	0.369	0.417
IIC	0.547	0.455	0.484	0.483	0.368	0.402
SSC-AI-RG-1	0.523	0.465	0.478	0.499	0.241	0.312
SSC-AI-RG-3	0.497	0.459	0.464	0.469	0.243	0.307
LSIR-1	0.552	0.439	0.479	0.508	0.36	0.409
TREDENCE-1	0.45	0.47	0.45	0.347	0.412	0.362
TREDENCE-2	0.477	0.437	0.448	0.394	0.37	0.368
TREDENCE-3	0.477	0.437	0.448	0.394	0.37	0.368
SSC-AI-RG-2	0.457	0.457	0.444	0.408	0.247	0.293
MACQUARIE-1	0.471	0.413	0.431	0.384	0.35	0.352
MACQUARIE-3	0.467	0.41	0.428	0.384	0.347	0.351
MACQUARIE-2	0.466	0.408	0.427	0.381	0.345	0.349
LSIR-3	0.461	0.41	0.425	0.411	0.213	0.27
AO-LANCS	0.312	0.227	0.257	0.253	0.155	0.185

Table 7: ROUGE-L and ROUGE-SU4 on English dataset ordered by ROUGE-L F1 score

Appendix B - Greek Task Results

Model	R-1 / R	R-1 / P	R-1 / F	R-2 / R	R-2 / P	R-2 / F
SSC-AI-RG-3	0.34	0.442	0.381	0.14	0.296	0.185
SSC-AI-RG-1	0.34	0.442	0.381	0.14	0.296	0.185
SSC-AI-RG-2	0.34	0.442	0.381	0.14	0.296	0.185
LSIR-1	0.297	0.421	0.346	0.112	0.203	0.141
TREDENCE-1	0.154	0.574	0.234	0.097	0.321	0.138
TREDENCE-2	0.154	0.574	0.234	0.097	0.321	0.138
AO-LANCS	0.284	0.448	0.344	0.091	0.276	0.131
LSIR-3	0.26	0.404	0.315	0.106	0.177	0.13
LSIR-2	0.246	0.42	0.309	0.089	0.174	0.115
LSIR-4	0.248	0.418	0.309	0.09	0.169	0.115
IIC	0.215	0.473	0.294	0.063	0.215	0.095
TREDENCE-3	0.068	0.683	0.119	0.043	0.415	0.072
LIPI	0.101	0.625	0.17	0.026	0.33	0.046

Table 8: ROUGE-1 and ROUGE-2 on Greek dataset ordered by R2 F1 score

Model	R-L / R	R-L / P	R-L / F	R-SU4 / R	R-SU4 / P	R-SU4 / F
SSC-AI-RG-3	0.247	0.348	0.284	0.177	0.328	0.226
SSC-AI-RG-1	0.247	0.348	0.284	0.177	0.328	0.226
SSC-AI-RG-2	0.247	0.348	0.284	0.177	0.328	0.226
LSIR-1	0.234	0.319	0.267	0.151	0.253	0.186
AO-LANCS	0.208	0.341	0.252	0.134	0.31	0.182
LSIR-3	0.205	0.293	0.238	0.145	0.202	0.167
LSIR-4	0.185	0.299	0.225	0.134	0.205	0.16
LSIR-2	0.183	0.3	0.224	0.132	0.207	0.159
IIC	0.165	0.353	0.222	0.106	0.247	0.146
TREDENCE-1	0.138	0.641	0.217	0.105	0.351	0.15
TREDENCE-2	0.138	0.641	0.217	0.105	0.351	0.15
TREDENCE-3	0.084	0.672	0.144	0.046	0.439	0.077
LIPI	0.081	0.509	0.137	0.046	0.402	0.08

Table 9: ROUGE-L and ROUGE-SU4 on Greek dataset ordered by ROUGE-L F1 score

Appendix C - Spanish Task Results

Model	R-1 / R	R-1 / P	R-1 / F	R-2 / R	R-2 / P	R-2 / F
LSIR-1	0.54	0.425	0.466	0.177	0.147	0.157
SSC-AI-RG-3	0.505	0.419	0.449	0.167	0.136	0.146
SSC-AI-RG-1	0.505	0.419	0.449	0.167	0.136	0.146
SSC-AI-RG-2	0.505	0.419	0.449	0.167	0.136	0.146
LSIR-3	0.511	0.429	0.454	0.158	0.129	0.138
AO-LANCS	0.503	0.425	0.448	0.15	0.128	0.134
TREDENCE-2	0.445	0.506	0.438	0.134	0.149	0.131
TREDENCE-1	0.445	0.506	0.438	0.134	0.149	0.131
TREDENCE-3	0.445	0.506	0.438	0.134	0.149	0.131
LSIR-2	0.497	0.418	0.443	0.149	0.122	0.131
LSIR-4	0.501	0.421	0.449	0.144	0.118	0.128
IIC	0.396	0.488	0.407	0.122	0.155	0.125
LIPI	0.142	0.58	0.217	0.045	0.196	0.07

Table 10: ROUGE-1 and ROUGE-2 on Spanish dataset ordered by R2 F1 score

Model	R-L / R	R-L / P	R-L / F	R-SU4 / R	R-SU4 / P	R-SU4 / F
LSIR-1	0.259	0.226	0.238	0.264	0.222	0.236
TREDENCE-2	0.192	0.238	0.2	0.212	0.24	0.208
TREDENCE-1	0.192	0.238	0.2	0.212	0.24	0.208
TREDENCE-3	0.192	0.238	0.2	0.212	0.24	0.208
LSIR-3	0.183	0.162	0.168	0.249	0.201	0.217
SSC-AI-RG-3	0.178	0.167	0.168	0.25	0.201	0.218
SSC-AI-RG-1	0.178	0.167	0.168	0.25	0.201	0.218
SSC-AI-RG-2	0.178	0.167	0.168	0.25	0.201	0.218
LSIR-2	0.178	0.163	0.167	0.241	0.195	0.21
AO-LANCS	0.194	0.147	0.164	0.238	0.199	0.211
IIC	0.143	0.204	0.159	0.194	0.236	0.197
LSIR-4	0.171	0.152	0.159	0.238	0.192	0.209
LIPI	0.098	0.325	0.146	0.069	0.291	0.107

Table 11: ROUGE-L and ROUGE-SU4 on Spanish dataset ordered by ROUGE-L F1 score