

University of Manchester  
School of Computer Science  
Project Report 2023

**Extractive Summarisation of  
UK Annual Reports**

Author: Vladislav Yotkov

Supervisor: Dr. Jonathan Shapiro

## **Abstract**

### **Extractive Summarisation of UK Annual Reports**

**Author: Vladislav Yotkov**

Although there has been considerable progress in Natural Language Processing (NLP) over the years, it has not been quite integrated in the Accounting and Finance (AF) industry. In the meantime, companies worldwide produce vast amounts of textual data as part of their reporting packages to comply with regulations and inform shareholders of their financial performance. The glossy annual report is such an example, widely read by investors but also quite long. Inspired by the Financial Narrative Summarisation (FNS) 2021 Task, we will design an Automatic Text Summarisation (ATS) system for the narrative parts of UK financial annual reports. With this goal in mind, we will implement and explore the following models for Extractive Text Summarisation (ETS): 1. custom Recurrent Neural Network (RNN), 2. fine-tuned FinBERT. In terms of evaluation, we will use the ROUGE metric to compare the performance of these models against standard ATS baselines: TextRank, and LexRank.

**Supervisor: Dr. Jonathan Shapiro**

# Contents

0.1	Introduction . . . . .	4
0.1.1	Financial Reports . . . . .	4
0.1.2	UK annual reports . . . . .	4
0.1.3	NLP in Accounting and Finance . . . . .	5
	<b>Bibliography</b>	<b>7</b>

# List of Figures

# List of Tables

## 0.1 Introduction

### 0.1.1 Financial Reports

Due to international regulations, companies are obliged to report their periodic performance (annual, bi-annual, quarterly) to various regulatory authorities<sup>1</sup> and other users (e.g., corporate stakeholders, investors, customers, suppliers, etc.). These reports contain essential information about the operations and finances of a business and are crucial for making informed decisions (from user perspective), but differ in regulatory forms. For example,

1. 10-K reports filed to the SEC<sup>2</sup> and accessible through their Electronic Data Gathering, Analysis, and Retrieval<sup>3</sup> (EDGAR) system are only for US registered businesses. They follow a standardised template and are plain text, which makes them particularly easy for automated large-scale research ([EHAR<sup>+</sup>19]). Also, the contents of these reports is quite strict, requiring solely five information sections<sup>4</sup>.
2. UK annual reports, the regulation of which is overseen by UK's Financial Reporting Council (FRC) are typically the main annual reporting method (also provided as PDF files). Unlike the 10-K, they are glossy and more stakeholder-oriented, and enjoy unlimited discretion over non-mandated content ([EHAR<sup>+</sup>19]) (e.g., photography and company brand material, non-mandatory narrative sections, etc.). However, these are more challenging for automated processing, due to their variable section structure, formatting and rich visual representations.

### 0.1.2 UK annual reports

The annual report is a key corporate disclosure, which is legally required for publicly companies by regulatory authorities. While it *does not have a rigid document structure* like the 10-K, it typically has a *narrative component*<sup>5</sup> and the financial statements (at the rear).

---

<sup>1</sup>Regulation authorities worldwide:

- Securities and Exchange Commission (SEC) in the USA
- European Securities and Markets Authority (ESMA) in Europe
- Financial Reporting Council (FRC) in the UK
- International Financial Reporting Standards (IFRS) in 167 jurisdictions worldwide

<sup>2</sup><https://www.sec.gov>

<sup>3</sup><https://www.sec.gov/edgar>

<sup>4</sup>(a) Business Overview (b) Risk Factors (c) Management's Discussion and Analysis of Financial Condition and Results of Operations (MD&A) (d) Financial Statements (e) Supplementary Disclosures

<sup>5</sup>The narrative component of a UK annual report typically consists of 1. Management's Commentary 2. Letter to Shareholders 3. Corporate Governance Statement 4. Auditor's Report 5. Remuneration Report

As we outlined in the Section 0.1.1, UK annual reports have the following inconvenient properties with regards to large-scale text understanding.

- They are very long documents. Throughout the years, their average length has been increasing significantly with the number of pages rising 57% for the median report from 2003 to 2016 (47 to 74 pages, respectively) ([LY19]), due to additional regulations between 2006 and 2008 ([EHAR<sup>+</sup>19]).
- They have variable nomenclature. From firm to firm, naming conventions vary “dramatically”, with more than 20 unique titles for various sections (e.g., Chair’s letter to shareholders, Management Commentary) ([LY19]).
- They incorporate embedded info-graphics. While domain experts hail the integration of highly interactive elements into corporate reporting ([KB16]), the compilation to PDF makes the task of analysing such unstructured documents automatically even harder ([LY19]).

These challenges motivate the work of [EHRY<sup>+</sup>19] who (a) established a set of 8 generic section headers<sup>6</sup> and (b) built the CFIE-FRSE<sup>7</sup> extraction tool for annual reports in three languages: English, Spanish and Portuguese.

As outlined in [Eil98], investors’ trust in the accountability of businesses would be based no longer as much on just the financial statements, but also on more descriptive narratives that define strategy and planning of resource use.

### 0.1.3 NLP in Accounting and Finance

The relevance of this project should also be understood from the perspective of the development of Natural Language Processing (NLP) in the Accounting and Finance (AF) domain. While some recognise the importance of understanding in-domain textual information ([L<sup>+</sup>10]), others like [EHRW<sup>+</sup>19] report that the industry is still doubtful and cynical about the NLP applications in the analysis of financial market disclosures. Furthermore, the latter also observe that AF researchers rely extensively on bag-of-words models, which are *not sufficient to encode complex contextual and semantic meaning* (especially in a domain with such *specialized language*). As for ATS [CHW19], is said to be the single AF study into disclosure summarisation and it demonstrates that machine-generated summaries are less likely to positively bias investors’ decisions compared to managerial ones.

---

6. Business Review 7. Environmental, Social, and Governance (ESG) Report 8. Risk Management Report

<sup>6</sup>(a) Chairman Statement (b) CEO Review (c) Corporate Governance Report (d) Directors Remuneration Report (e) Business Review (f) Financial Review (g) Operating Review (h) Highlights

<sup>7</sup>The CFIE-FRSE stands for Corporate Financial Information Environment - Final Report Structure Extractor It is publicly available at <https://github.com/drelhaj/CFIE-FRSE>.

This only confirms that there is a wide gap of NLP applications in Accounting research, and this further motivates our work.



# Bibliography

- [CHW19] Eddy Cardinaels, Stephan Hollander, and Brian White. Automatic summarization of earnings releases: attributes and effects on investors' judgments. *Review of Accounting Studies*, 24, 09 2019.
- [EHAR<sup>+</sup>19] Mahmoud El-Haj, Paulo Alves, Paul Rayson, Martin Walker, and Steven Young. Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as pdf files. *Accounting and Business Research*, 2019. Forthcoming.
- [EHRW<sup>+</sup>19] Mahmoud El-Haj, Paul Rayson, Martin Walker, Steven Young, and Vasiliki Simaki. 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, 2019.
- [EHRY<sup>+</sup>19] Mahmoud El Haj, Paul Edward Rayson, Steven Eric Young, Paulo Alves, and Carlos Herrero Zorita. *Multilingual Financial Narrative Processing: Analysing Annual Reports in English, Spanish and Portuguese*. World Scientific Publishing, 2019.
- [Ell98] Robert K Elliott. Accounting in the 21st century. 1998.
- [KB16] P Kriz and H Blomme. The future of corporate reporting—creating the dynamics for change. *International Federation of Accountants (IFAC)*, available at: [www.ifac.org/global-knowledgegateway/viewpoints/future-corporate-reporting-creating-dynamics-change](http://www.ifac.org/global-knowledgegateway/viewpoints/future-corporate-reporting-creating-dynamics-change) (accessed 29 May 2016), 2016.
- [L<sup>+</sup>10] Feng Li et al. Textual analysis of corporate disclosures: A survey of the literature. *Journal of accounting literature*, 29(1):143–165, 2010.
- [LY19] Craig Lewis and Steven Young. Fad or future? automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, 49(5):587–615, 2019.