

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

0.1 Introduction

0.1.1 Financial Reports

Due to international regulations, companies are obliged to report their periodic performance (annual, bi-annual) to various regulatory authorities¹ and other users (e.g., corporate stakeholders, investors, customers, suppliers, etc.). These reports contain essential information about the operations and finances of the business, but differ in forms. For example,

1. 10-K reports filed to the SEC² and accessible through their Electronic Data Gathering, Analysis, and Retrieval³ (EDGAR) system. They follow a standardised template and are plain text, which makes them particularly easy for automated large-scale research ([EHAR⁺19]).

As outlined in [Ell98], 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.

¹Regulation 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

²<https://www.sec.gov>

³<https://www.sec.gov/edgar>

Bibliography

- [EHAR⁺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.
- [Eli98] Robert K Elliott. Accounting in the 21st century. 1998.