



# Financial Narrative Processing: a Corpus-based NLP approach

Mahmoud El-Haj  
CASS, Lancaster University



# Research Projects

CorpComm

- ESRC

CFIE

- ESRC

SenseSourcing

- ESRC/UCREL

VardSourcing

- Jisc

Arabic Dialects

- UCREL

Strategic Contents

- ICRA/UCREL/LUMS

# Financial Narrative Projects



# Project Team

CorpComm <sup>1</sup>

CFIE <sup>2</sup>

CASS

Prof Tony McEnergy<sup>1</sup>

Dr Mahmoud El-Haj<sup>1,2</sup>

Dr Andrew Hardie<sup>1</sup>

SCC

Dr Paul Rayson<sup>1,2</sup>

LUMS

Prof Steve Young<sup>1,2</sup>

MBS

Prof Martin Walker<sup>1,2</sup>

Dr Thomas Schleicher<sup>2</sup>

LSE

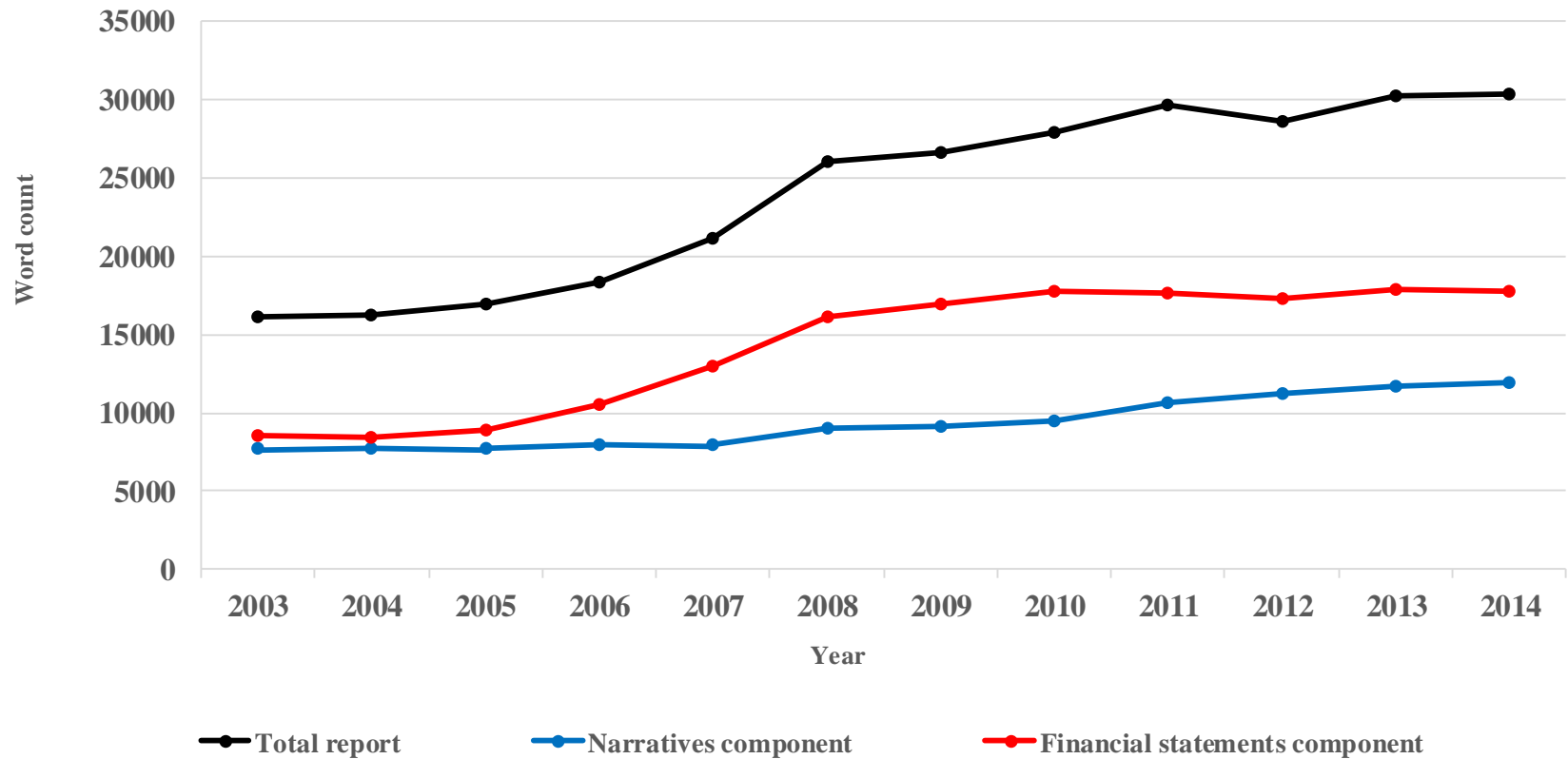
Dr Vasiliki Athanasakou<sup>2</sup>

# What is it about?

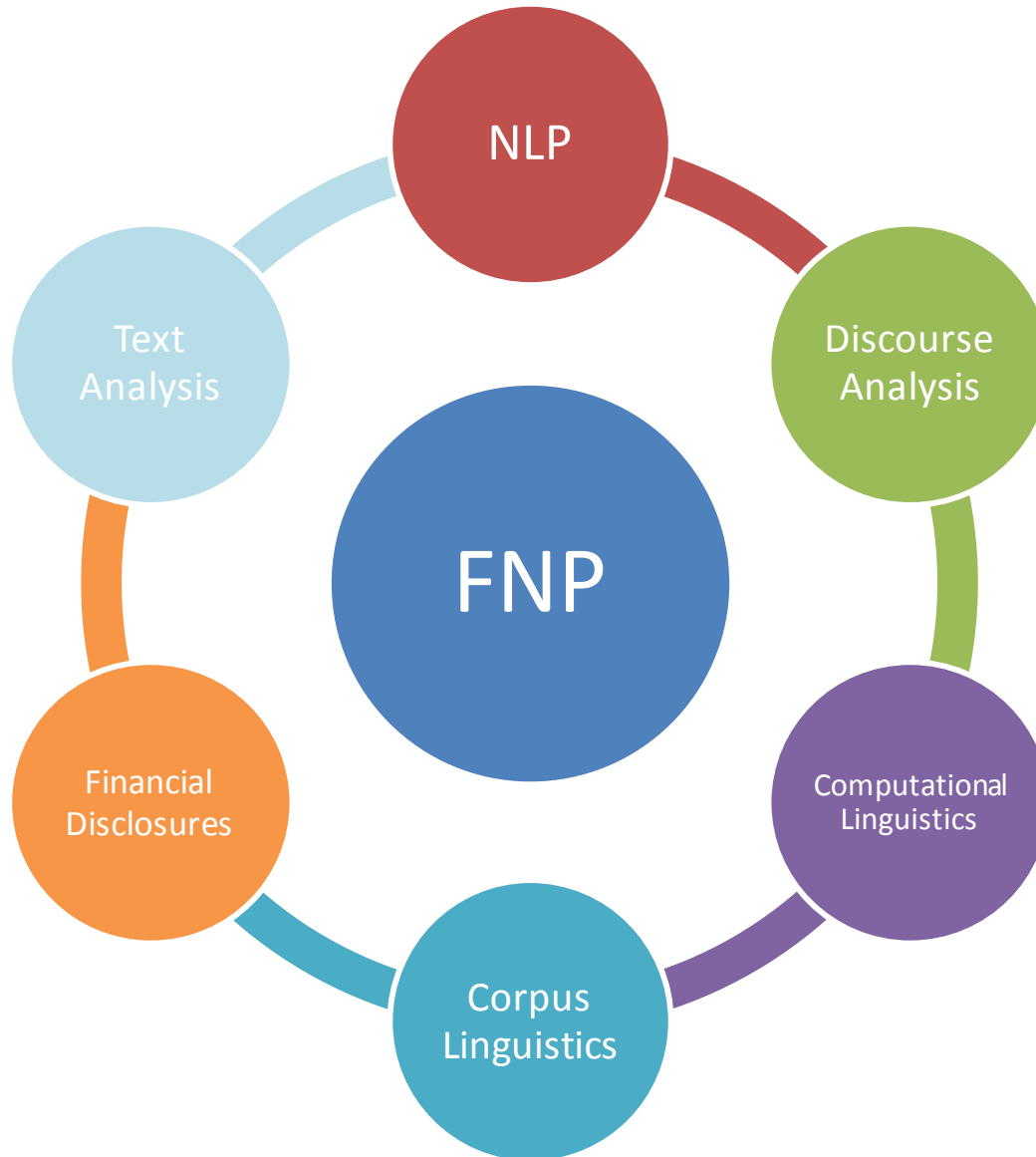
Project analyses U.K. financial narratives, their association with financial statement information, and their informativeness for investors.

# Why Financial Narratives

a doubling of the median word count over the sample period



# FNP: Fields of Study

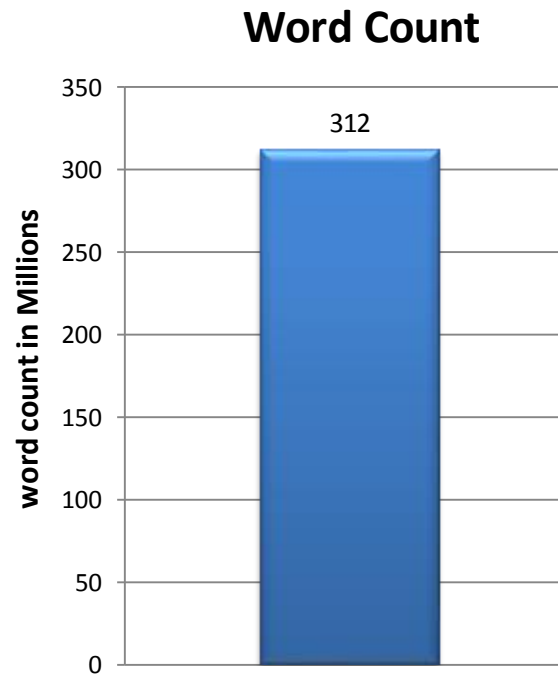
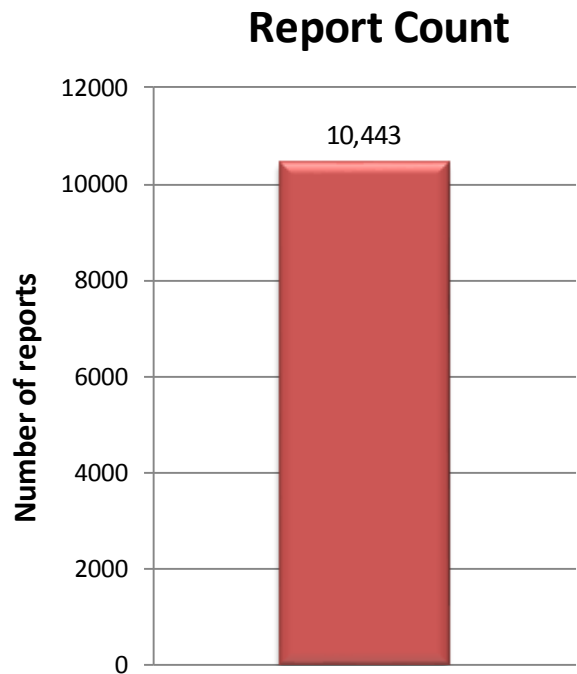


# Big Data

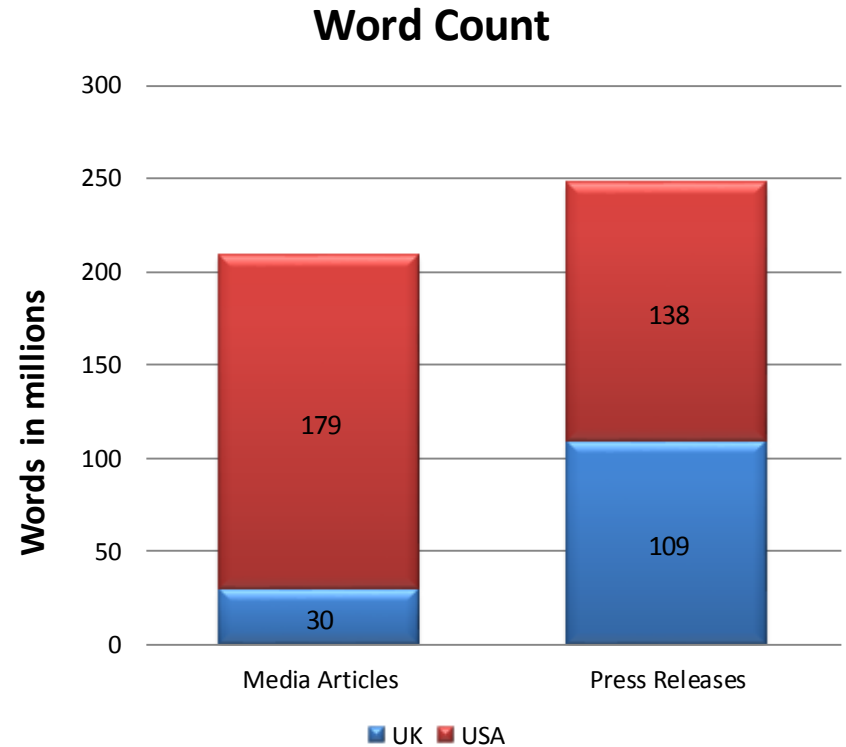
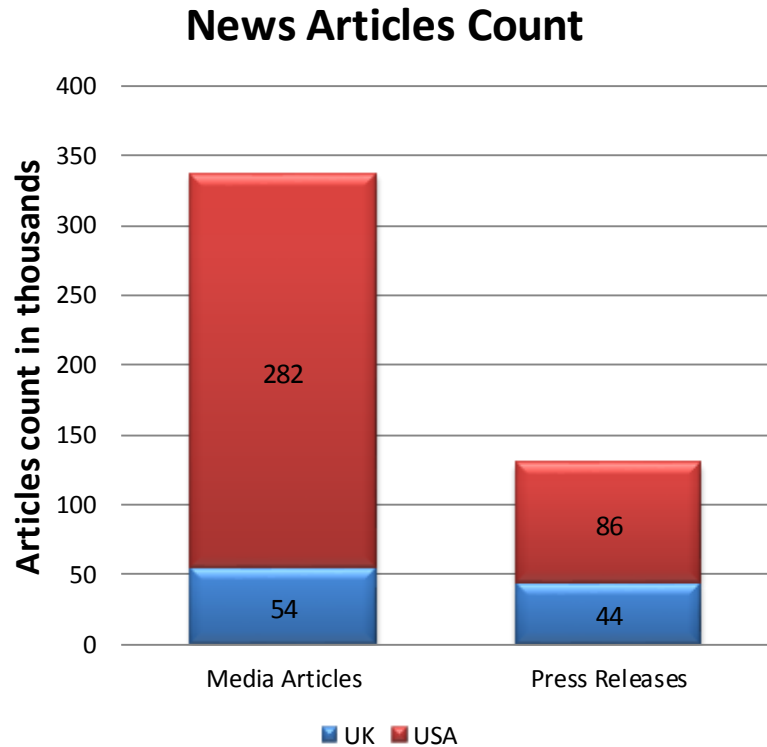




# Annual Reports: how big?



# Financial Media: how big?

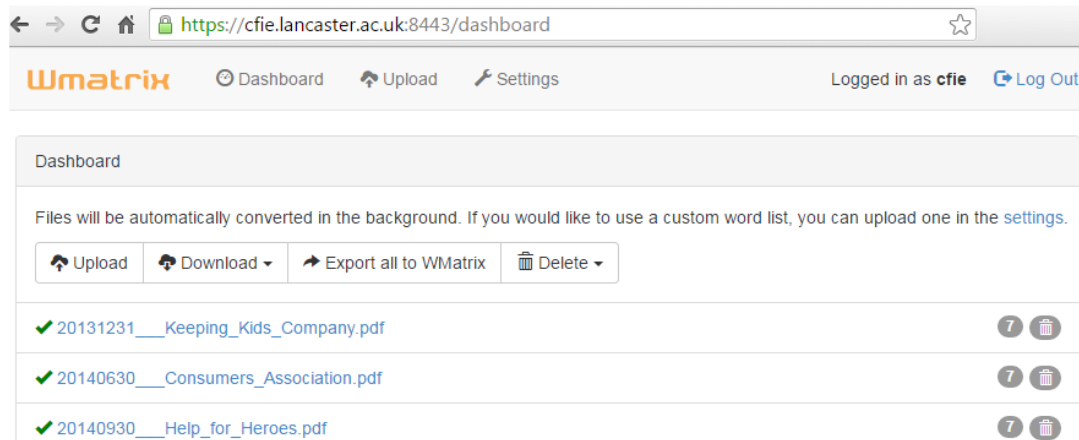
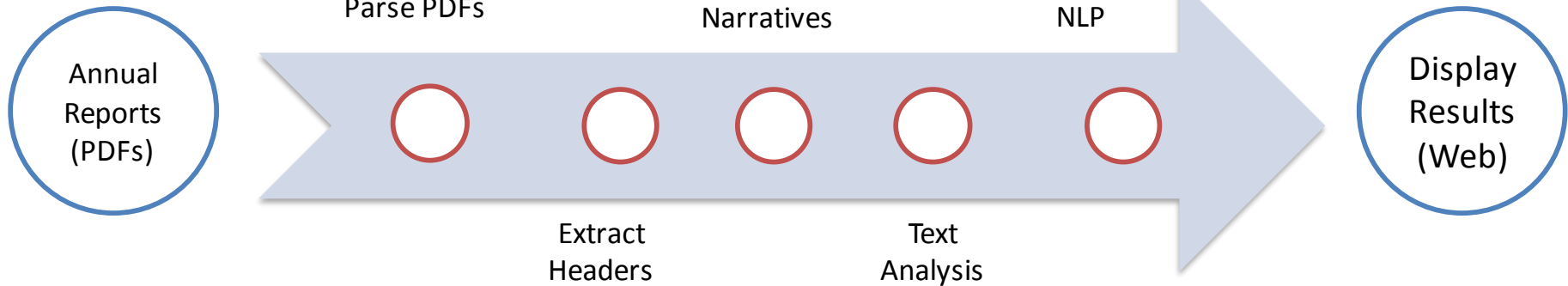


# Output

- NLP publicly available tool ([Wmatrix-Import](#))
- First large scale study of UK annual reports structure.
- First ever published disclosure scores for UK annual reports.
- Approaches used help speed up the analysis process and close the gap between firms and investors.
- Leads to better understand corporate financial decisions and corporate financial performance.

Wmatrix

# Wmatrix-Import



# UK Annual Reports Sample



## Financial highlights

### Sales

**+6.8%**

Sales (including VAT, including fuel)

### Underlying operating profit

**£789m**

Underlying operating profit up 6.9%

### Underlying profit before tax

**£712m**

Underlying profit before tax up 7.1%

### Return on capital employed

**11.1%**

Return on capital employed

### Underlying basic earnings

**28.1p**

Underlying basic earnings per share up 6.0%

## Contents

### Business review

Financial highlights	2
Chairman's letter	4
Chief Executive's letter	6
Market overview	8
Key performance indicators	10
Our strategy	12
Great food	14
Compelling general merchandise & clothing	16
Complementary channels & services	18
Developing new business	20
Growing space & creating property value	22



## Content

Spirax Sarco at a glance	1
Chairman's statement	2
Business review	3
Market overview	4
Performance review	5
Board of Directors	6
Directors' report	7
Corporate governance	8
Corporate social responsibility	9
The Directors' remuneration report	10
Statement of Directors' responsibilities	11
Financial statements	12
Report of the independent auditor	13
Group income statement	14
Balance sheets	15
Statements of recognised income	16
Cash flow statements	17
Notes to the accounts	18
Financial summary	19
Officers and advisers	20

Comments: 03 Chairman's report 04 Business review 13 Finance commentary 18 Report of the Directors 19 Board of meeting 20 Profit and loss account 20 Statement of total recognised gains and losses 21 Balance sheet 22 Cash flow statements 23 Statement of accounting policies 25 Notes to the accounts 27 Corporate governance 33 Directors' responsibilities in respect of the accounts 35 Report of the Independent Auditor 40 Company information

### Financial highlights

Turnover: £54,570 +7.2%	Operating profit: £3,277 +2.8%	Dividend	Debt levels

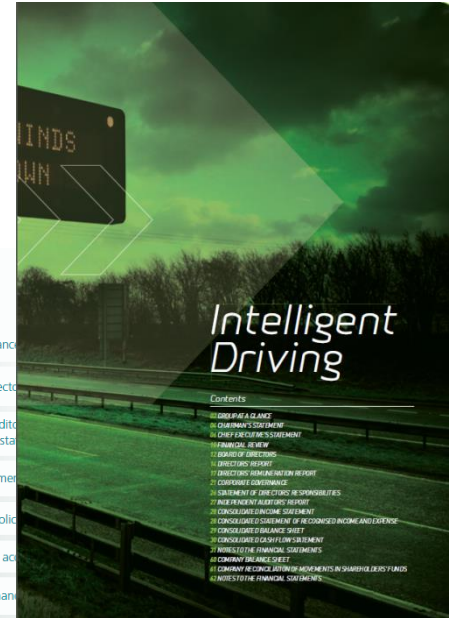
### Carbon report

Carbon resulting from fuel used for combustion, owned transport, process emissions (Tonnes of CO <sub>2</sub> )	Carbon emissions from purchased electricity (Tonnes of CO <sub>2</sub> )
2,324	1,060
2011: 2,284	2011: 1,080
2010: 2,283	2010: 1,181

Carbon emissions from purchased electricity (Tonnes of CO<sub>2</sub>)

0.0184	0.0077
2011: 0.0184	2011: 0.0077
2010: 0.0190	2010: 0.0076

02 Who we are and what we do	42 Corporate governance
06 24 hours in the life of Arriva	46 Statement of directors' responsibilities
08 Our growth story	47 Independent auditor's report
10 Our markets	48 Financial statements
12 Chairman's statement	52 Accounting policies
14 Chief executive's review	56 Notes to the accounts
22 Financial review	82 Five-year financial performance
26 Corporate responsibility	83 Parent company financial statements
32 Board of directors	90 Statement of directors' responsibilities on the parent company financial statements
34 Directors' report	91 Independent auditors' report on the parent company financial statements
37 Directors' remuneration report	92 Financial calendar, registered office and advisors



## Intelligent Driving

### Contents

- 01 ARRIVA AT A GLANCE
- 02 CHAIRMAN'S STATEMENT
- 03 CHIEF EXECUTIVE'S REVIEW
- 04 FINANCIAL REVIEW
- 05 BOARD OF DIRECTORS
- 06 DIRECTORS' REMUNERATION REPORT
- 07 INDEPENDENT AUDITOR'S REPORT
- 08 STATEMENT OF DIRECTORS' RESPONSIBILITIES
- 09 ARRIVAL AT ARRIVA'S REPORT
- 10 CONSOLIDATED STATEMENT OF RECOGNISED INCOME AND EXPENSE
- 11 CONSOLIDATED BALANCE SHEET
- 12 CONSOLIDATED CASH FLOW STATEMENT
- 13 PARENT COMPANY FINANCIAL STATEMENTS
- 14 COMPANY REPORTING REQUIREMENTS IN SHAREHOLDERS' FUNDS
- 15 NOTES TO THE FINANCIAL STATEMENTS

# UK Annual report tool: Extraction

- Steps in extraction process:
  - Detect contents page
  - Parse contents page
  - Extract section
  - Detect section type
  - Reorder section

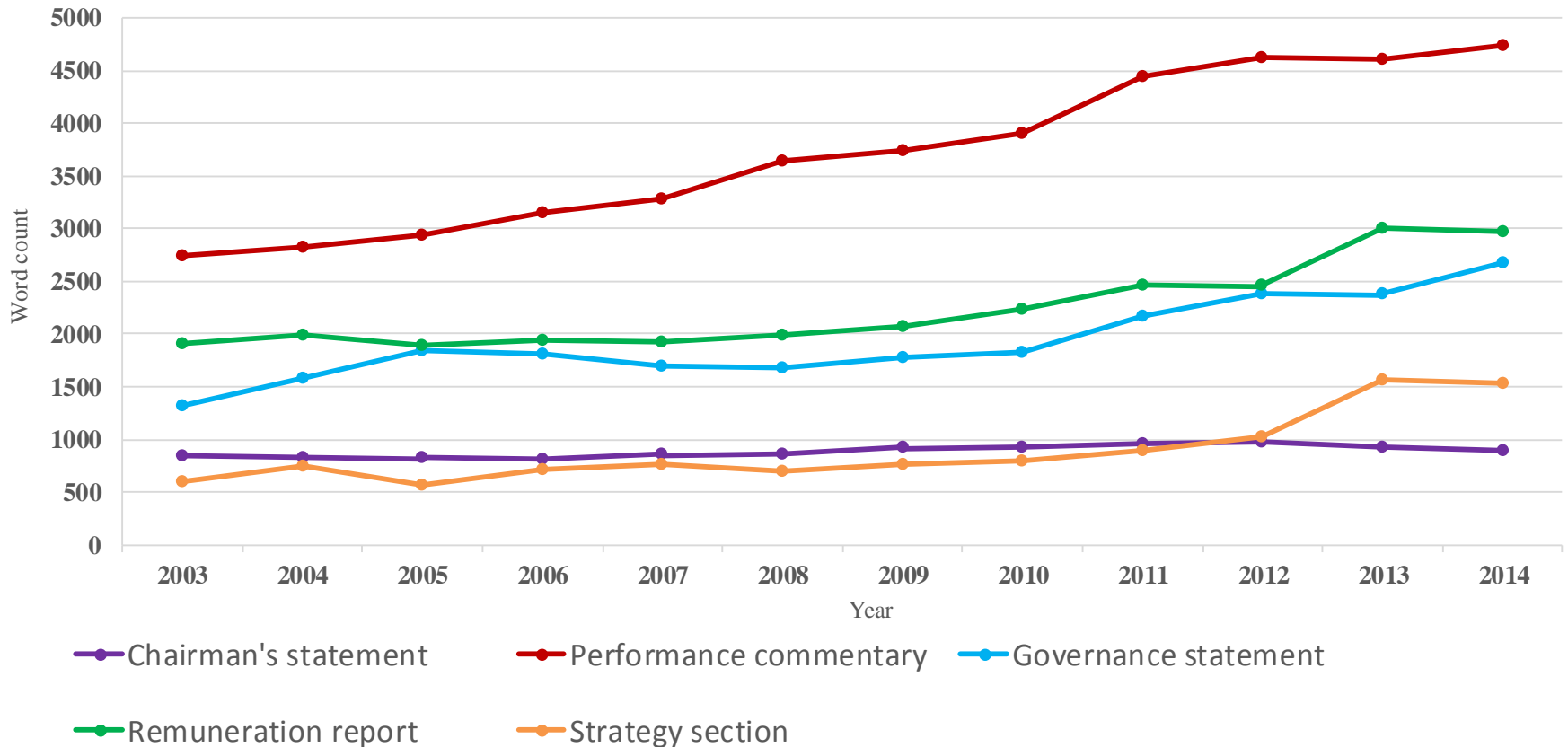


# Techniques

- Automatically detect 8 key headers
- Reorder headers to front and rear
- String matching, regex
- Levenshtein Distance / text similarity
- Machine learning to classify headers
- Crowdsourcing
- Text Reuse
- Hand crafted keyword lists by experts
- Bottom up and top down word lists:
  - Computer based approaches
  - By Linguists

# Growing over time

how some of the main types of content have grown over the sample period





# Collaboration

FRC	<ul style="list-style-type: none"><li>•Funding</li><li>•Help standardise UK annual reports</li></ul>
Financial Firm, London	<ul style="list-style-type: none"><li>•Fraud Detection</li><li>•Management Change</li><li>•Management vs Analysts</li></ul>
Financial Services Provided, London	<ul style="list-style-type: none"><li>•Interest in the tool</li><li>•Automatic Structuring</li><li>•Indexing Annual Reports</li></ul>
Big Four in Accounting	<ul style="list-style-type: none"><li>•Interest in the tool</li><li>•Provided free-to-publish annual reports</li></ul>
Big Data company, London	<ul style="list-style-type: none"><li>•Interest in working together</li></ul>
KSU, Saudi Arabia	<ul style="list-style-type: none"><li>•Arabic Financial Narratives (Book Chapter)</li></ul>
Sunway University, Malaysia	<ul style="list-style-type: none"><li>•Malay Financial Narratives (consultancy)</li></ul>
National Natural Science Foundation of China (NSFC)	<ul style="list-style-type: none"><li>•Chinese Financial Narratives (consultancy)</li></ul>

# Thanks

More about the projects: <http://ucrel.lancs.ac.uk/cfie/>

Wmatrix-Import: <https://cfie.lancaster.ac.uk:8443/>