



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 ¹

CFIE ²

CASS

Prof Tony McEnery¹

Dr Mahmoud El-Haj^{1,2}

Dr Andrew Hardie¹

SCC

Dr Paul Rayson^{1,2}

LUMS

Prof Steve Young^{1,2}

MBS

Prof Martin Walker^{1,2}

Dr Thomas Schleicher²

LSE

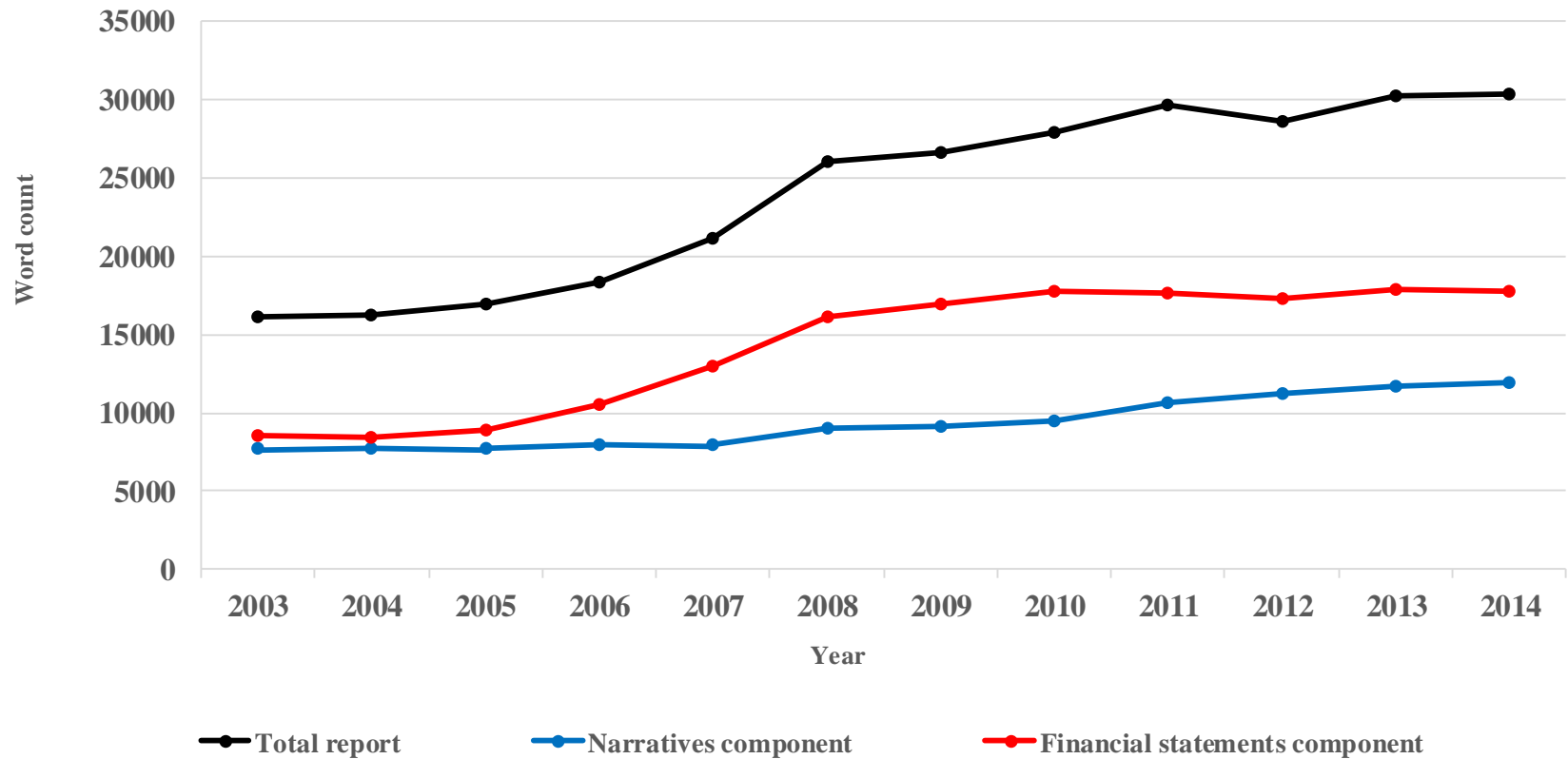
Dr Vasiliki Athanasakou²

What is it about?

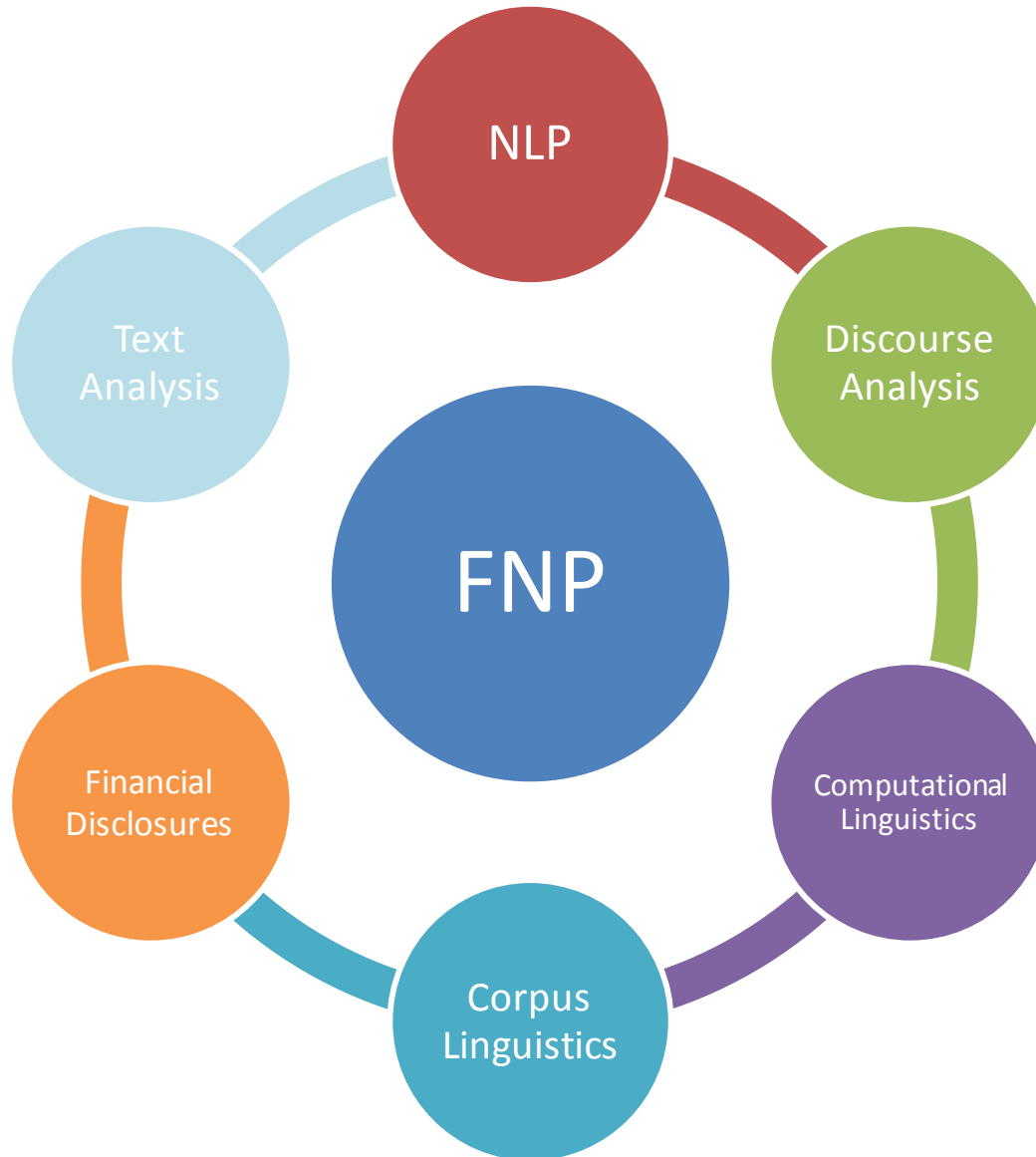
Project analyses U.K. financial narratives, their association with financial statement information, and their informativeness for investors.

Why Financial Narratives

a doubling of the median word count over the sample period



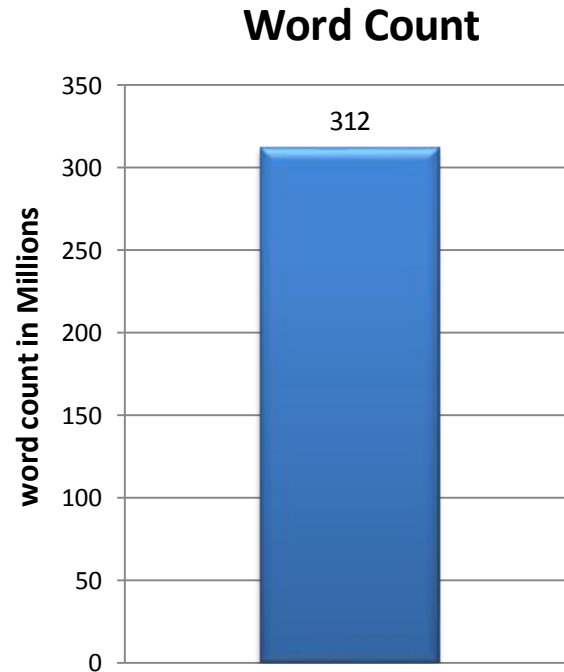
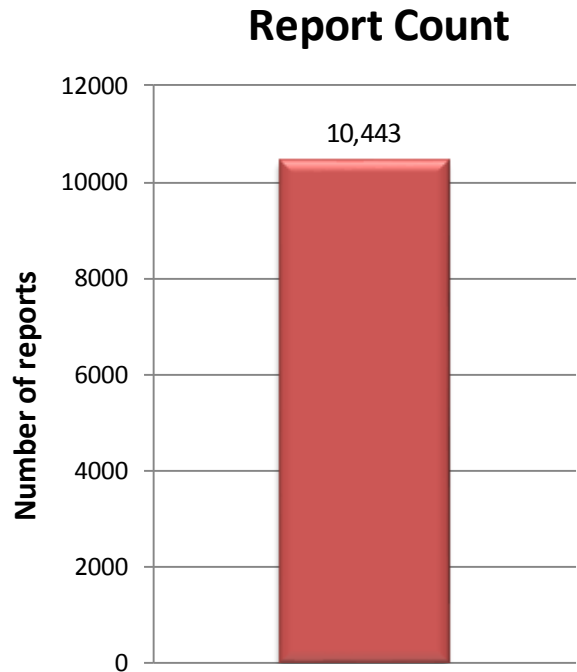
FNP: Fields of Study



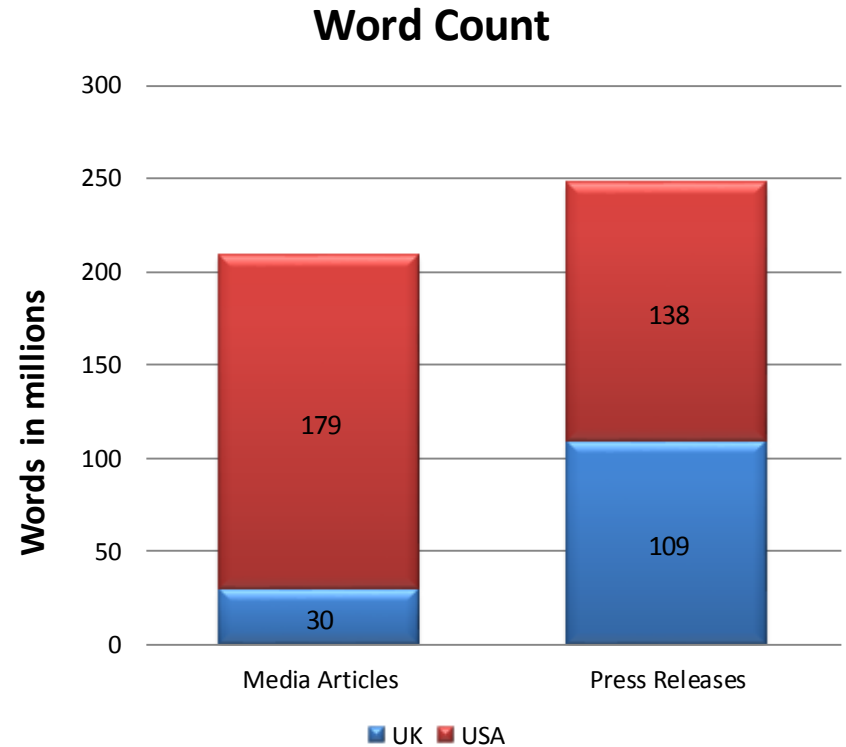
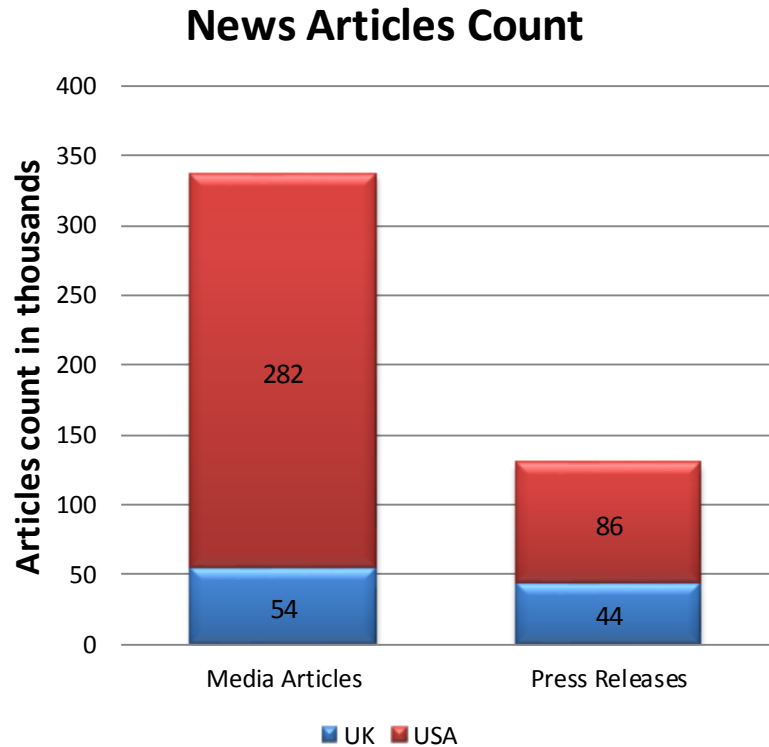
Big Data



Annual Reports: how big?



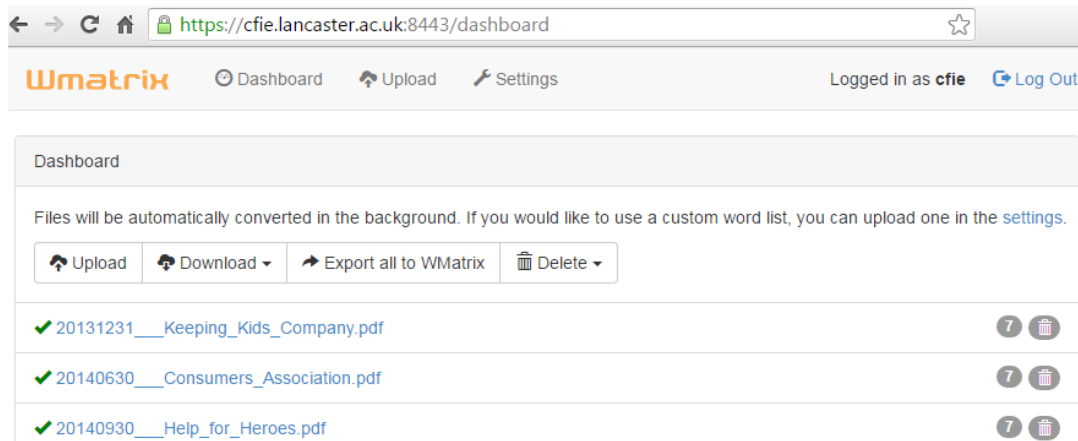
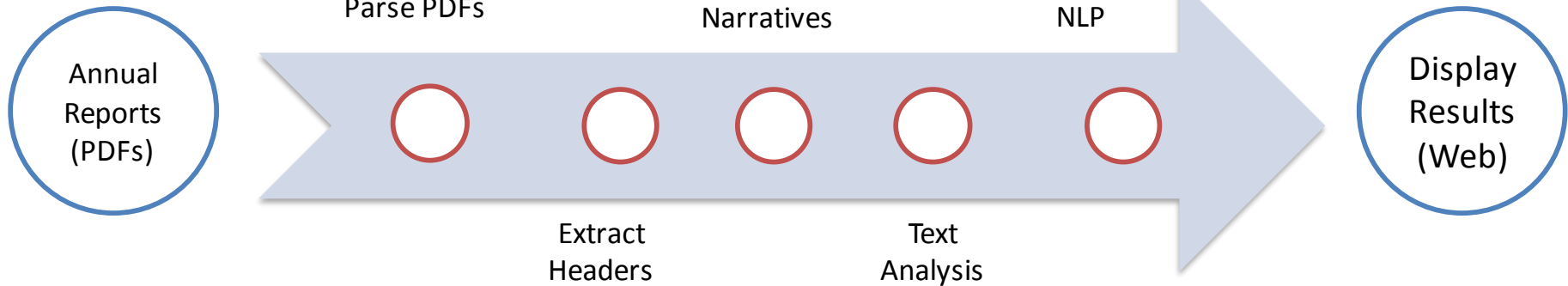
Financial Media: how big?



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-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
Chairman's letter
Chief Executive's letter
Market overview
Key performance indicators
Our strategy
Great food
Competing general merchandise & clothing
Complementary channels & services
Developing new business
Growing space & creating property value



Content

- Spirax Sarco at a glance.....
- Chairman's statement.....
- Business review.....
- Market overview.....
- Performance review.....
- Board of Directors.....
- Directors' report.....
- Corporate governance.....
- Corporate social responsibility.....
- The Directors' remuneration report.....
- Statement of Directors' responsibilities.....
- Financial statements.....
- Report of the independent auditor.....
- Group income statement.....
- Balance sheets.....
- Statements of recognised income and expense.....
- Cash flow statements.....
- Notes to the accounts.....
- Financial summary.....
- Officers and advisers.....

Comments: 03 Chairman's report 04 Business review 13 Finance commentary 18 Report of the Directors 19 House 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 23 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: \uparrow £54,570 +7.2%

Operating profit: \uparrow £3,277 +2.8%

Dividend: \uparrow

Debt levels: \uparrow

Carbon report

Carbon resulting from fuel used for combustion, owned transport, process emissions (Tonnes of CO₂)

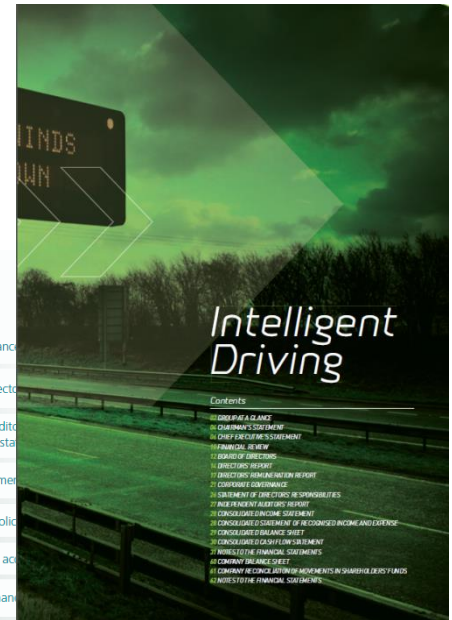
Carbon emissions from purchased electricity (Tonnes of CO₂)

2,324 1,060

0.0184 0.0077

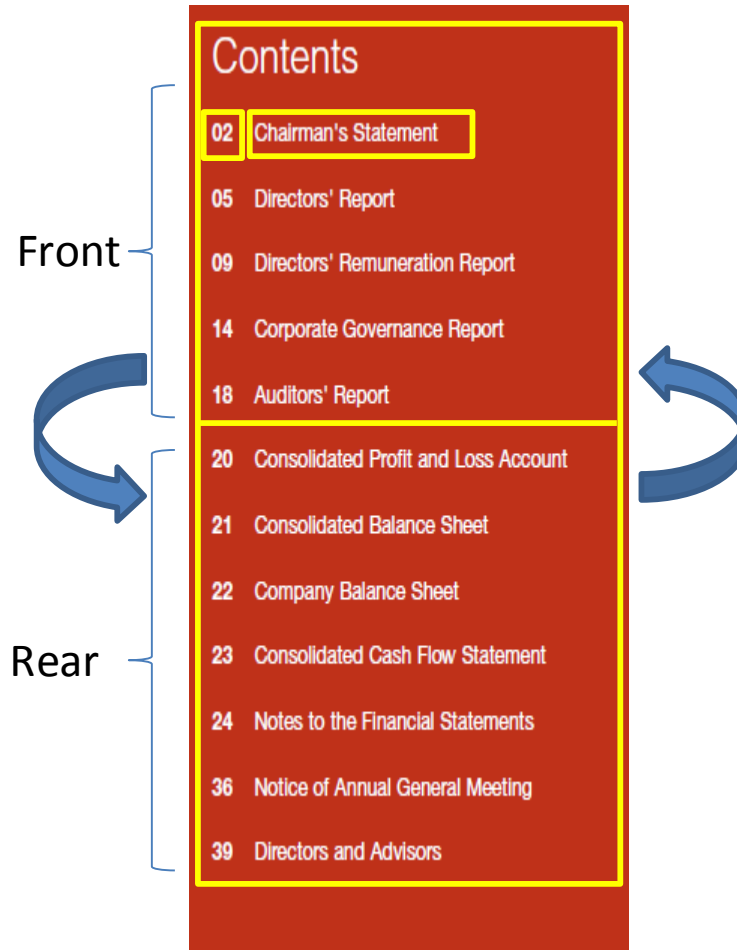
ARRIVA

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



UK Annual report tool: Extraction

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

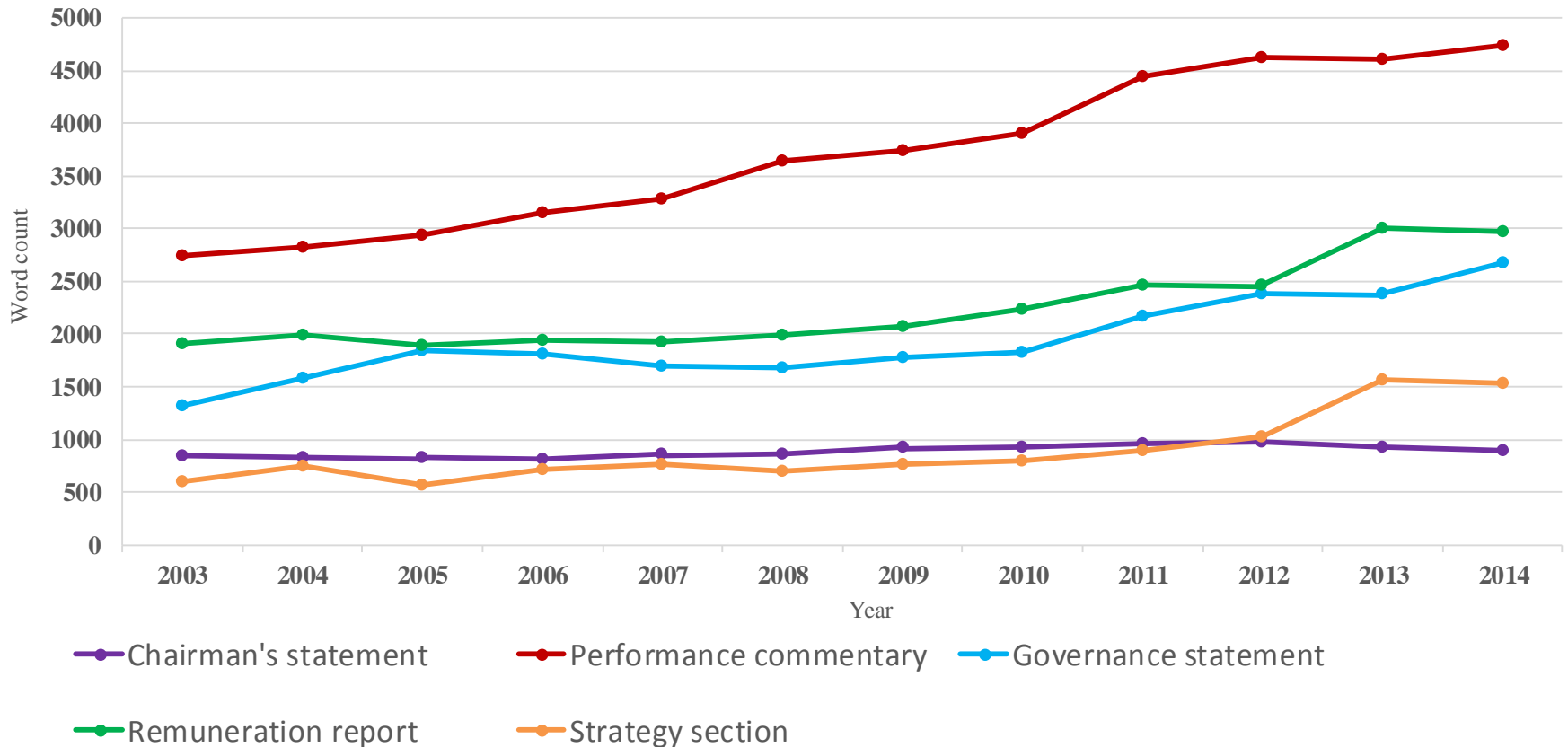


Techniques

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

Thanks

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

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