Financial **N**arrative **P**rocessing (FNP)

Mahmoud El-Haj

Lancaster University

DocElhaj

http:// www.lancaster.ac.uk/staff/elhaj/

Financial Narrative Projects

Project Team

CFIE [1]

CorpComm [2]

FNP [3]

SCC

Dr Mahmoud El-Haj[1,2,3]

Dr Paul Rayson[1,2,3]

LUMS

Prof Steve Young[1,2,3]

MBS

Prof Martin Walker[1,2,3]

Dr Thomas Schleicher[1]

CASS

Prof Tony McEnery[2]

Dr Andrew Hardie[2]

LSE

Dr Vasiliki Athanasakou[1]

What is it about?

The projects analyse U.K. financial narratives, their association with financial statement information, and their informativeness for investors.

Contributions

- Develop an automated method for extracting narrative content and structure from UK annual reports provided as PDFs

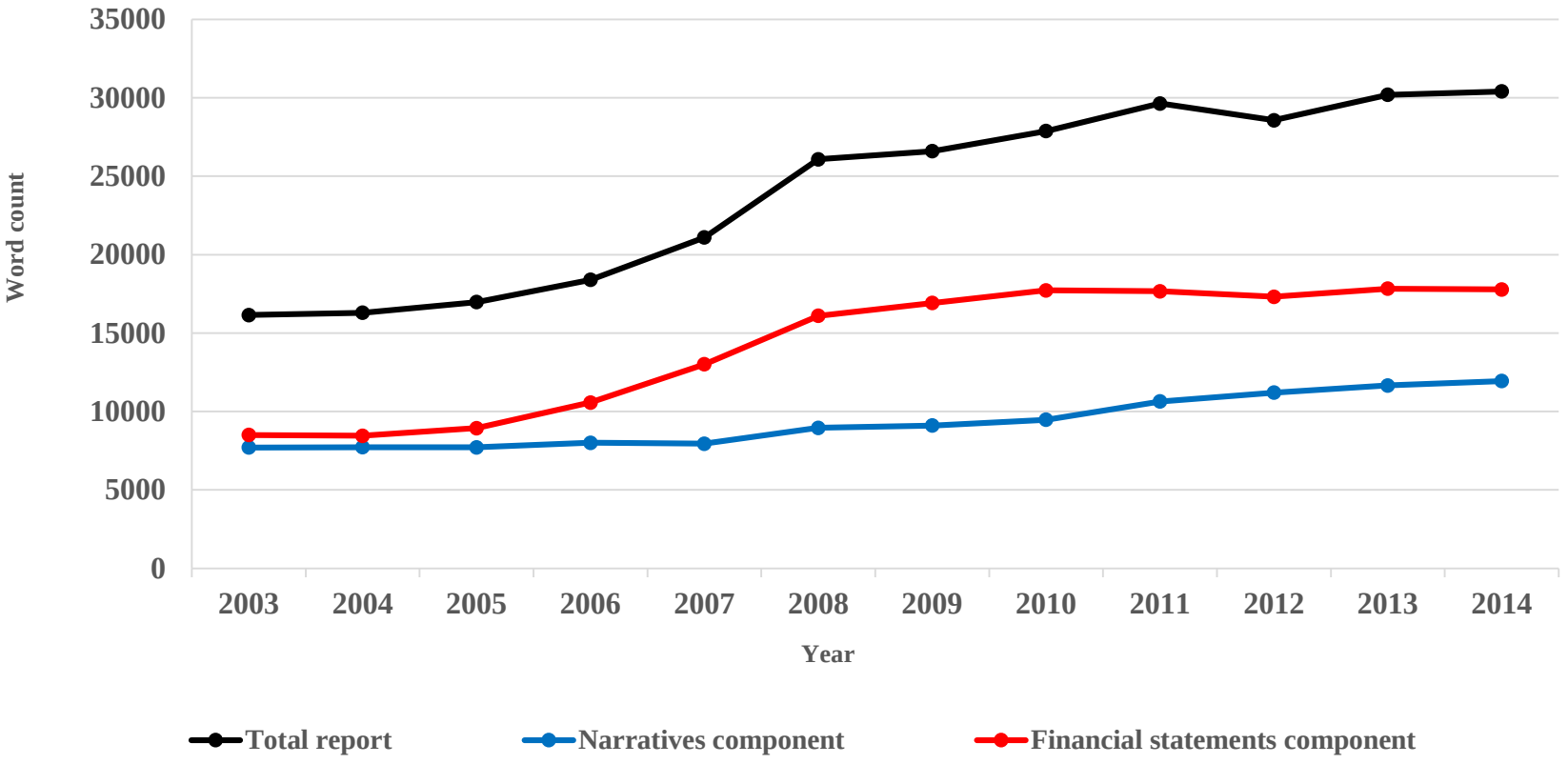
- First large-sample tests of the incremental predictive ability of UK annual report narratives

- First study to examine the incremental and differential predictive ability of alternative annual report narrative sections

- First study to model disagreement between preparers' and to examine the impact of disagreement on the predictive properties of narratives

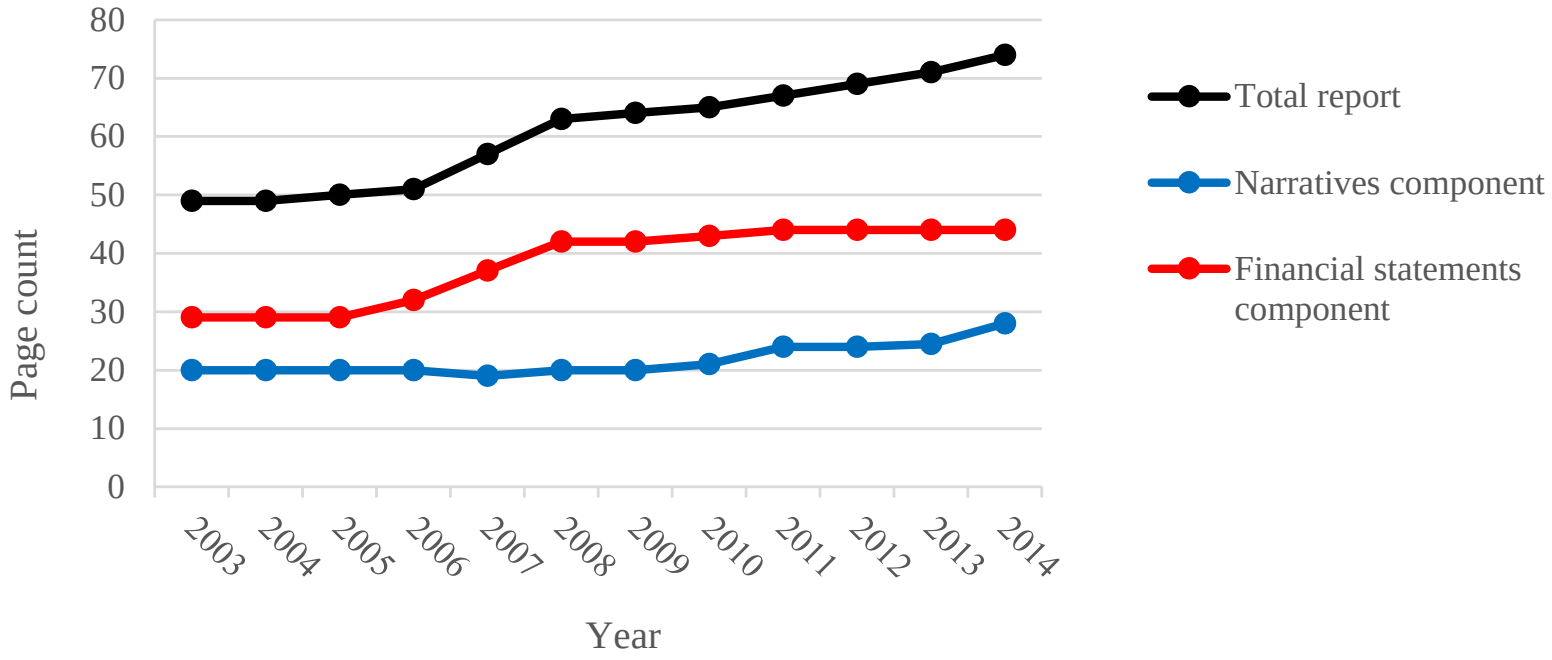
Why Financial Narratives

a doubling of the median word count over the sample period

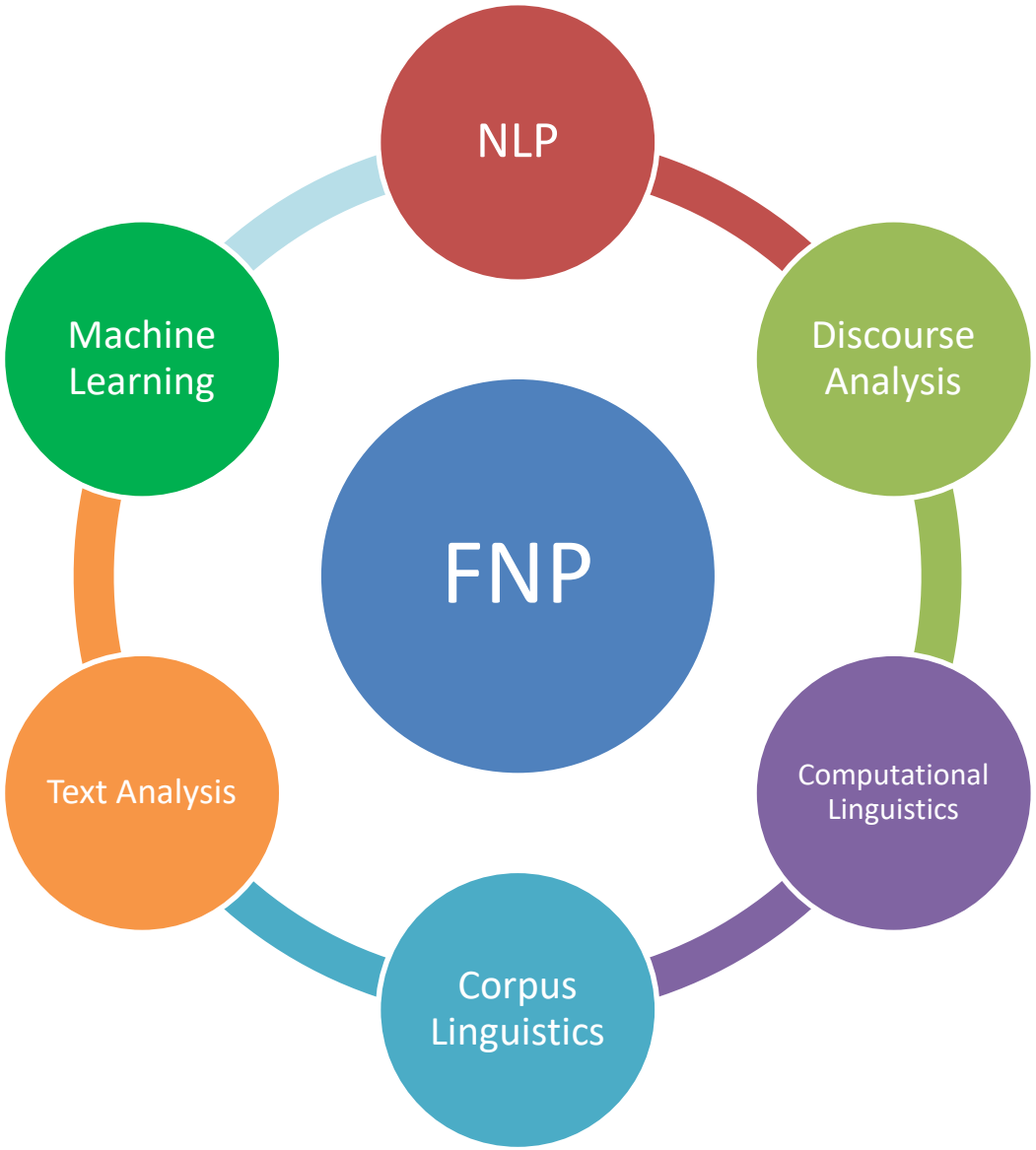


Why Financial Narratives

a doubling of the median page count over the sample period



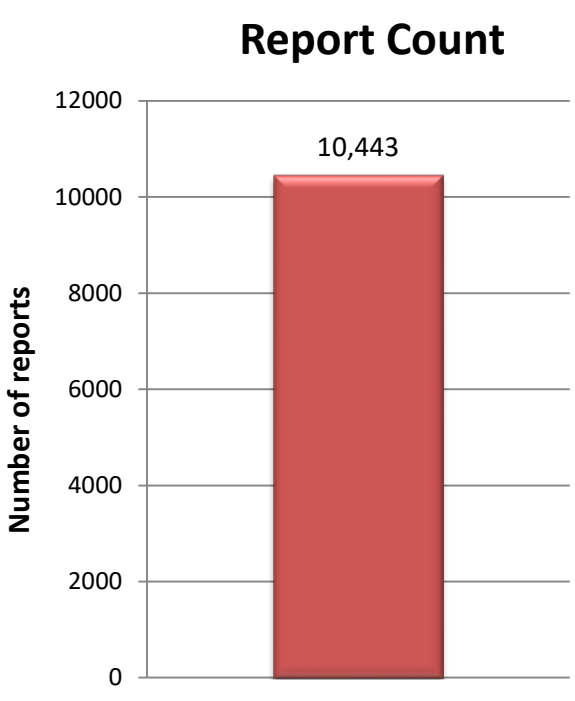
FNP: Fields of Study

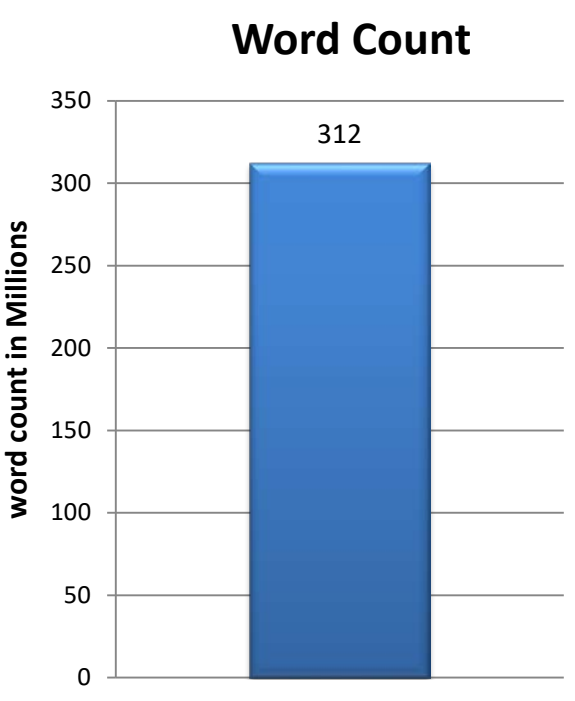


Big Data

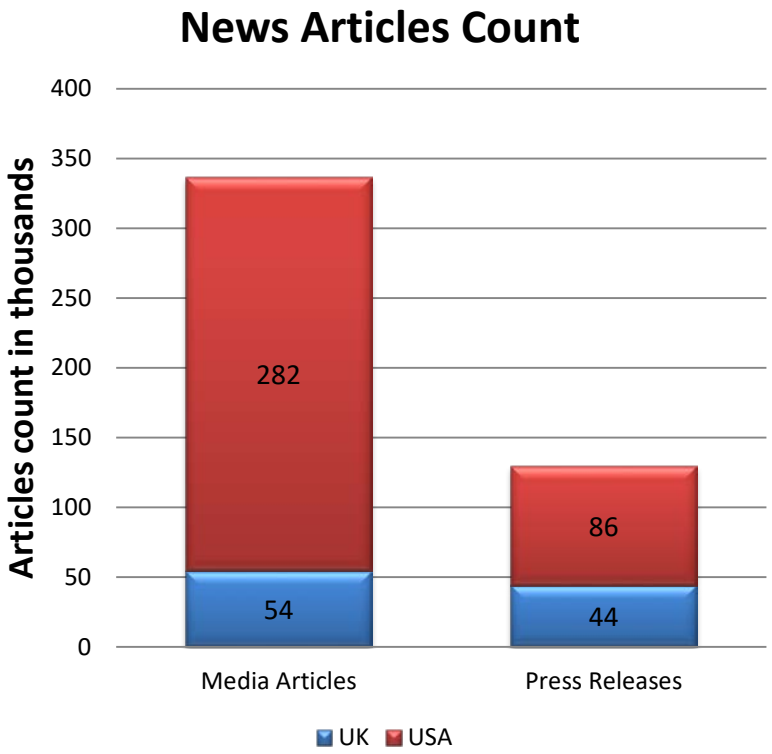
PDF
XML
JSON
DB
Financial Narratives
Annual Reports
PEAs
Conference Calls
Financial News
Press Releases
Media Articles
TMX
HTML
RTF
PLAIN

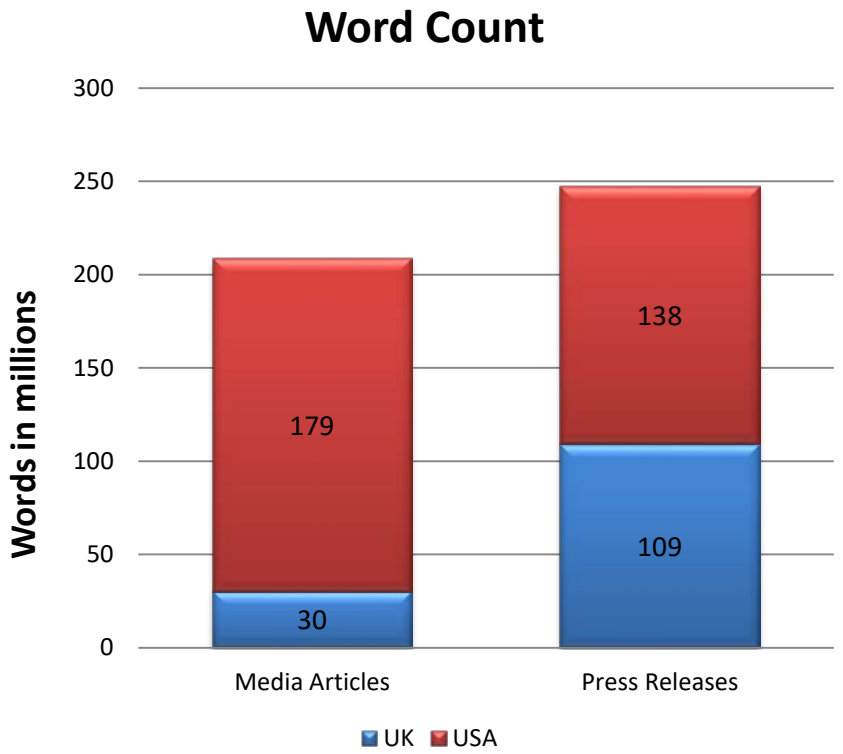
Annual Reports: how big?





Financial Media: how big?





Previous Work

Can we apply NLP methods used in prior US studies to UK annual reports?

US Filings

- US companies must submit:

1. 10-K: Annual
2. 10-Q: Quarterly
3. 8-K: Special Events (between 10-K and 10-Q)
4. Annual Report

10-K Annual Form

Each 10-K contains 4 parts and 15 items

- **PART I**
- **ITEM 1.** Description of Business
- **ITEM 2.** Description of Properties
- **ITEM 3.** Legal Proceedings
- **ITEM 4.** Mine Safety Disclosures
- **PART II**
- **ITEM 5.** Market for Registrant's Common Equity….
- **ITEM 6.** Selected Financial Data
- **ITEM 7.** Management's Discussion and Analysis….
- **ITEM 8.** Financial Statements and Supplementary Data
- **ITEM 9.** Changes in and Disagreements ….
- **PART III**
- **ITEM 10.** Directors, Executive Officers and Corporate Governance
- **ITEM 11.** Executive Compensation
- **ITEM 12.** Security Ownership of Certain Beneficial Owners….
- **ITEM 13.** Certain Relationships and Related Transactions….
- **ITEM 14.** Principal Accounting Fees and Services
- **PART IV**
- **ITEM 15.** Exhibits, Financial Statement Schedules….

10-K Annuals (Starbucks vs McDonald's)

Starbucks Corporation

	PART I
Item 1	Business
Item 1A	Risk Factors
Item 1B	Unresolved Staff Comments
Item 2	Properties
Item 3	Legal Proceedings
Item 4	*(Removed and Reserved)*
	PART II
Item 5	Market for the Registrant's C
Item 6	Selected Financial Data
Item 7	Management's Discussion an
Item 7A	Quantitative and Qualitative I
Item 8	Financial Statements and Sup
	Report of Independent Regist
Item 9	Changes in and Disagreemen
Item 9A	Controls and Procedures
Item 9B	Other Information
	PART III
Item 10	Directors, Executive Officers
Item 11	Executive Compensation

McDONALD'S CORPORATION

Part I.	
Item 1	Business
Item 1A	Risk Factors and Cautionary
Item 1B	Unresolved Staff Comments
Item 2	Properties
Item 3	Legal Proceedings
Item 4	Mine Safety Disclosures
Part II.	
Item 5	Market for Registrant's Comn
Item 6	Selected Financial Data
Item 7	Management's Discussion ar
Item 7A	Quantitative and Qualitative I
Item 8	Financial Statements and Su
Item 9	Changes in and Disagreemer
Item 9A	Controls and Procedures
Item 9B	Other Information
Part III.	
Item 10	Directors, Executive Officers
Item 11	Executive Compensation

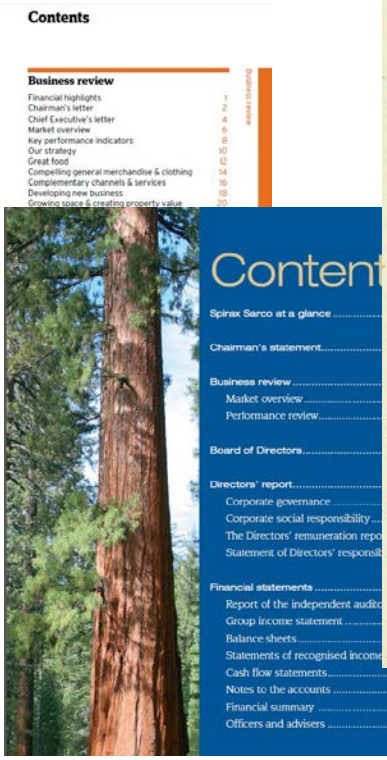
UK Annual Reports

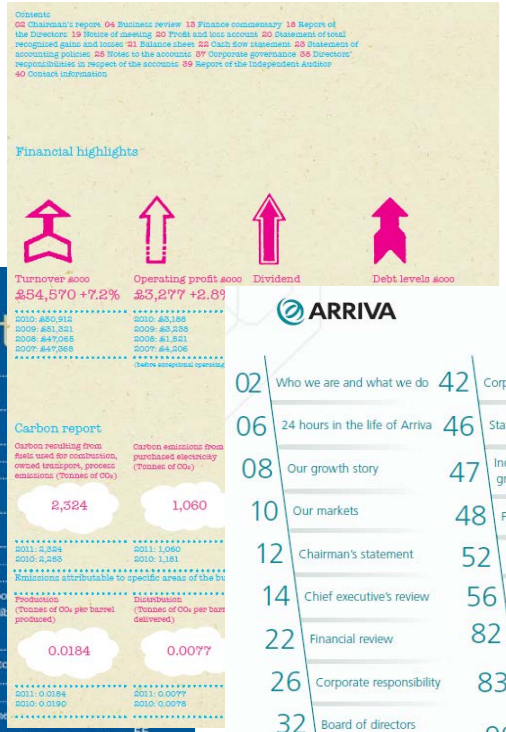
- Free style (no standard structure)
- Use of images, text, hyperlinks, …etc.
- PDF format
- Content and structure varies across firms.
- Management have more discretion over what, where, and how much information on topics such as risk, strategy, performance, etc. is reported.

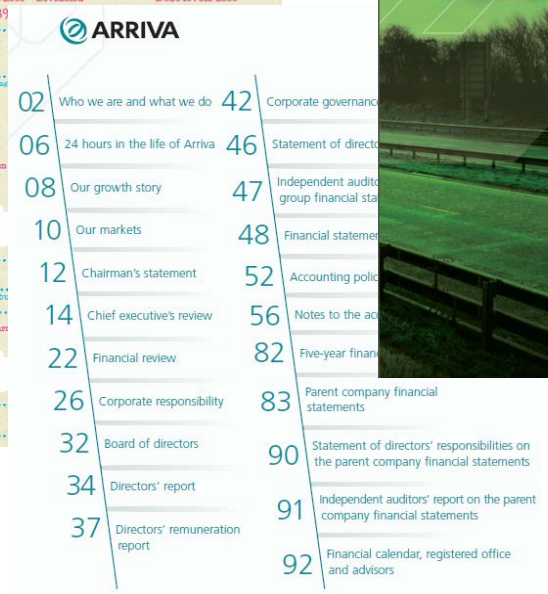
This makes the extraction and analysis task more challenging;

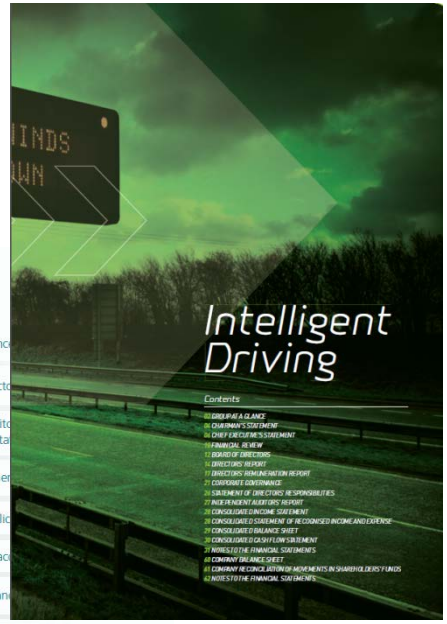
but it provides research opportunities.

UK Annual Reports Sample

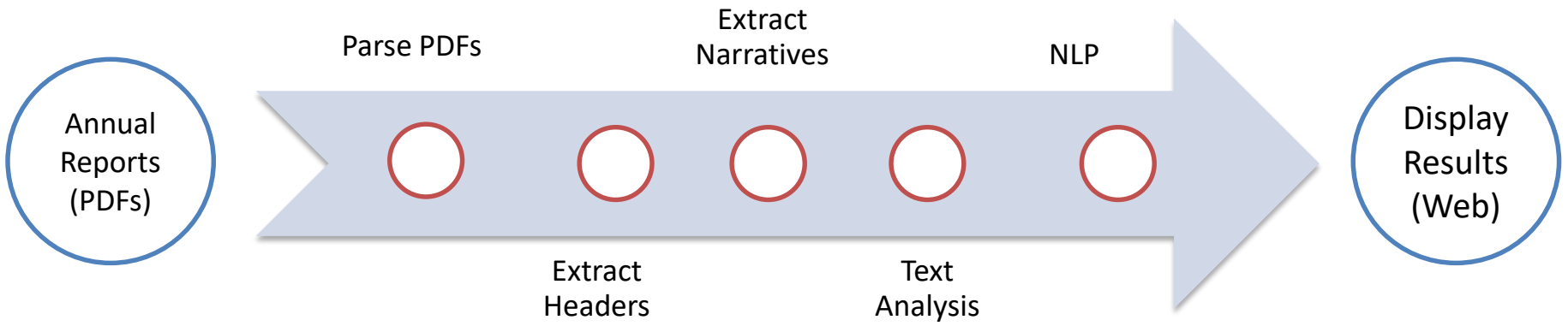


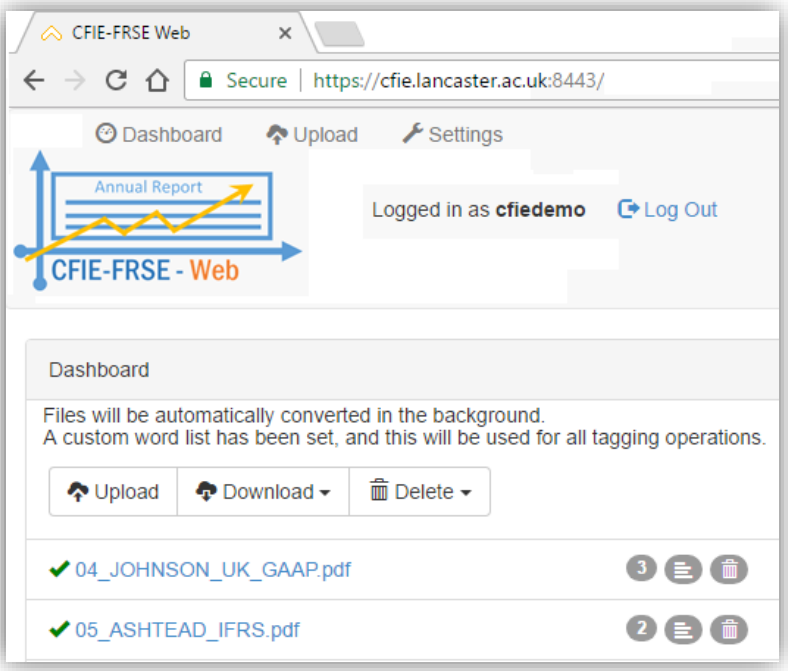






CFIE-FRSE Tool

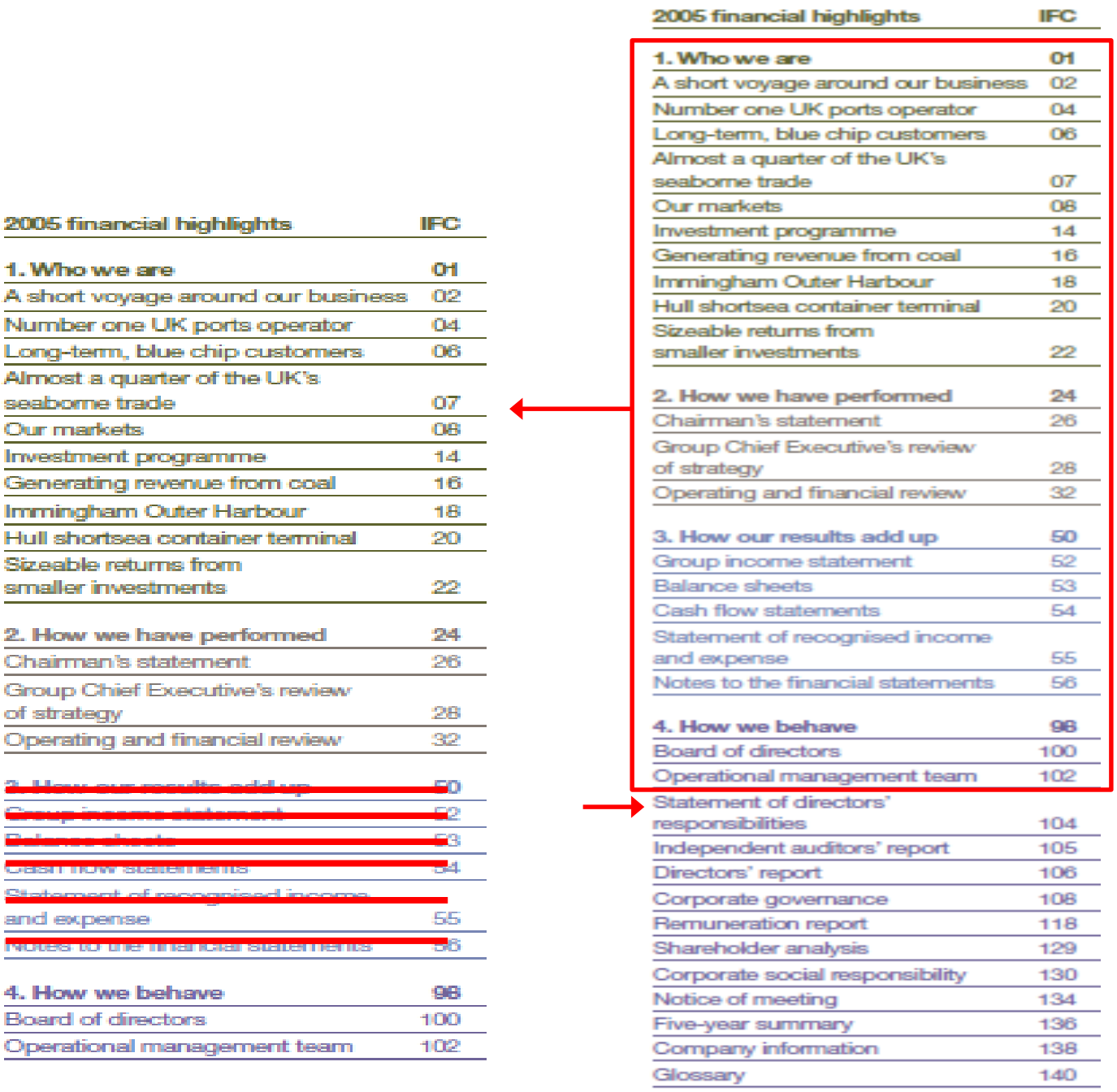




UK Annual report tool: Extraction

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

Example



2005 financial highlights IFC

1. Who we are 01
A short voyage around our business 02
Number one UK ports operator 04
Long-term, blue chip customers 06
Almost a quarter of the UK's seaborne trade 07
Our markets 08
Investment programme 14
Generating revenue from coal 16
Immingham Outer Harbour 18
Hull shortsea container terminal 20
Sizeable returns from smaller investments 22

2. How we have performed 24
Chairman's statement 26
Group Chief Executive's review of strategy 28
Operating and financial review 32

3. How our results add up 50
Group income statement 52
Balance sheets 53
Cash flow statements 54
Statement of recognised income and expense 55
Notes to the financial statements 56

4. How we behave 98
Board of directors 100
Operational management team 102

2005 financial highlights IFC

1. Who we are 01
A short voyage around our business 02
Number one UK ports operator 04
Long-term, blue chip customers 06
Almost a quarter of the UK's seaborne trade 07
Our markets 08
Investment programme 14
Generating revenue from coal 16
Immingham Outer Harbour 18
Hull shortsea container terminal 20
Sizeable returns from smaller investments 22

2. How we have performed 24
Chairman's statement 26
Group Chief Executive's review of strategy 28
Operating and financial review 32

3. How our results add up 50
Group income statement 52
Balance sheets 53
Cash flow statements 54
Statement of recognised income and expense 55
Notes to the financial statements 56

4. How we behave 98
Board of directors 100
Operational management team 102

Statement of directors' responsibilities 104
Independent auditors' report 105
Directors' report 106
Corporate governance 108
Remuneration report 118
Shareholder analysis 129
Corporate social responsibility 130
Notice of meeting 134
Five-year summary 136
Company information 138
Glossary 140

Example

2005 financial highlights	IFC
1. Who we are	**01**
A short voyage around our business	02
Number one UK ports operator	04
Long-term, blue chip customers	06
Almost a quarter of the UK's seaborne trade	07
Our markets	08
Investment programme	14
Generating revenue from coal	16
Immingham Outer Harbour	18
Hull shortsea container terminal	20
Sizeable returns from smaller investments	22
2. How we have performed	**24**
Chairman's statement	26
Group Chief Executive's review of strategy	28
Operating and financial review	32
4. How we behave	**98**
Board of directors	100
Operational management team	102
Directors' report	106
Corporate governance	108
Remuneration report	118
Corporate social responsibility	130

2005 financial highlights	IFC
1. Who we are	**01**
A short voyage around our business	02
Number one UK ports operator	04
Long-term, blue chip customers	06
Almost a quarter of the UK's seaborne trade	07
Our markets	08
Investment programme	14
Generating revenue from coal	16
Immingham Outer Harbour	18
Hull shortsea container terminal	20
Sizeable returns from smaller investments	22
2. How we have performed	**24**
Chairman's statement	26
Group Chief Executive's review of strategy	28
Operating and financial review	32
3. How our results add up	**50**
Group income statement	52
Balance sheets	53
Cash flow statements	54
Statement of recognised income and expense	55
Notes to the financial statements	56
4. How we behave	**98**
Board of directors	100
Operational management team	102
Statement of directors' responsibilities	104
Independent auditors' report	105
Directors' report	106
Corporate governance	108
Remuneration report	118
Shareholder analysis	129
Corporate social responsibility	130
Notice of meeting	134
Five-year summary	136
Company information	138
Glossary	140

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

Report Classification

- Heuristic Approach
- Machine Learning

- Classify narratives (front) component into the following generic categories to facilitate cross-sectional and temporal analysis:
  - Chairman's statement
  - Performance commentary (incl. CEO review, strategic review, finance director's review, operating review, business review, etc.)
  - Governance statement (incl. chairman's introduction, separate committee statements, statement on internal control, etc.)
  - Remuneration report
  - Residual commentary (incl. overview, highlights, CRS report, principal risks and uncertainties, directors' report, etc.)

Output

- NLP publicly available tool (CFIE-FRSE)
- 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.

SSRNPaper

- Heterogeneous Narrative Content in Annual Reports Published as PDF Files: Extraction, Classification and Incremental Predictive Ability

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2803275

Output Evaluation

- Manual evaluation
  - Precision (Type I errors) $\rightarrow$ false positives
  - Recall (Type II errors) $\rightarrow$ false negatives
  - Overall accuracy (F_1) $\rightarrow$ harmonic mean of precision and recall

- Compare extracted section headers with table of contents

- Assigned page numbers with actual page numbers from table of contents

- Examine accuracy of section classification

- Evaluations based on 586 reports selected at random (approx. 5% of initial population
  - 11,720 annual report sections

Output Evaluation

Panel A: Section extraction

			Error frequencies		Retrieval performance (%)		
	N actual	N extracted	Type 1	Type 2	Precision	Recall	F_1 score
Pooled annual report	11,009	10,820	286	475	97.47	95.69	96.57
Narratives component	5,237	5,233	216	220	96.04	95.80	95.92

Output Evaluation

Panel B: Page number synchronization

	Type I errors for section extraction treated as incorrect pagination			Type I errors for section extraction not treated as incorrect pagination		
	N	N errors	Precision (%)	N	N errors	Precision (%)
Pooled annual report	10,820	736	93.20	10,534	450	95.73
Narratives component	5,233	500	90.45	5,017	248	95.06

Output Evaluation

Panel C: Document classification

			Error frequencies		Retrieval performance (%)		
	N actual	N classified	Type 1	Type 2	Precision	Recall	F_1 score
Narratives component	4,929	4,846	88	83	98.22	98.32	98.27
Financials component	5,434	5,536	83	88	98.47	98.38	98.43

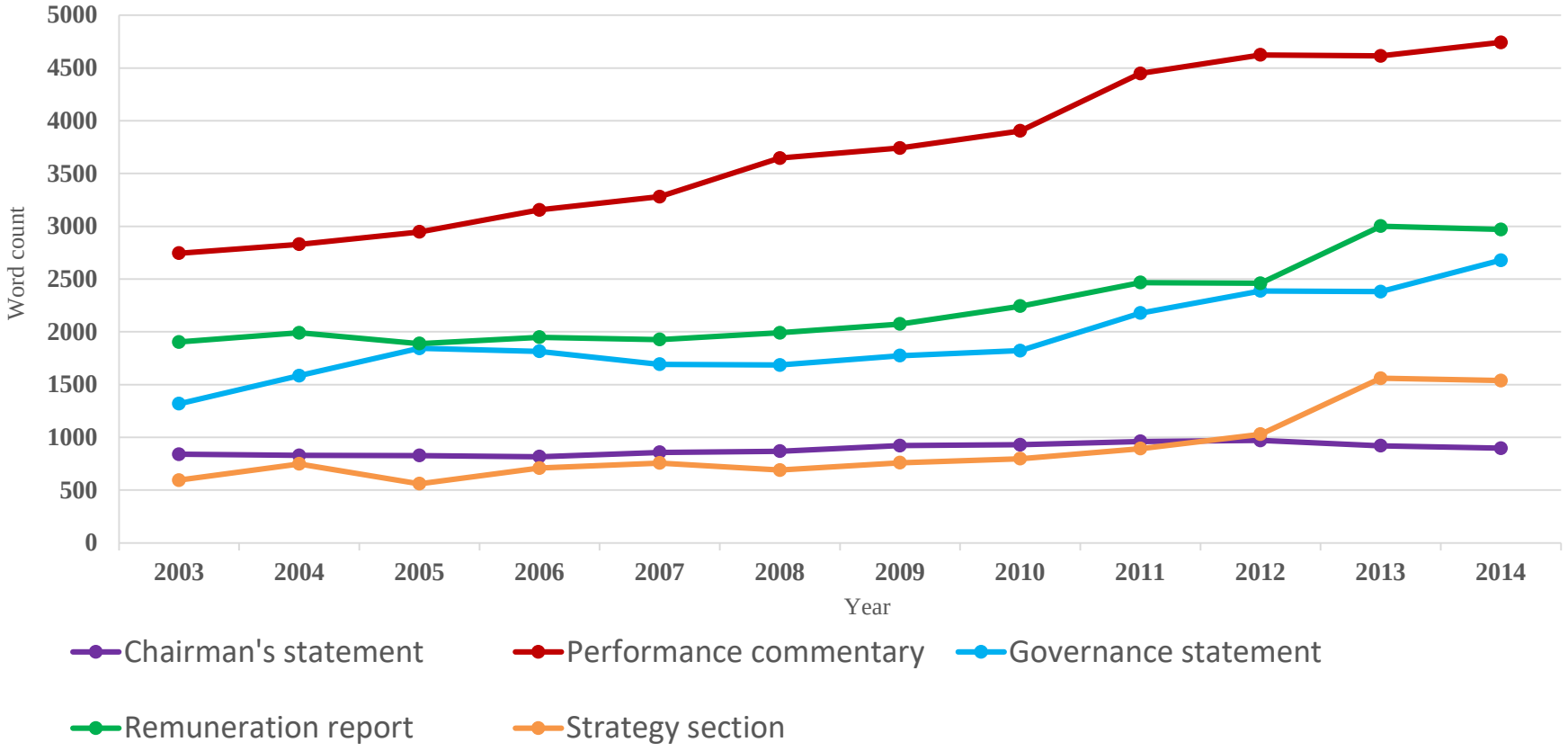
Output Evaluation

Panel C: Document classification

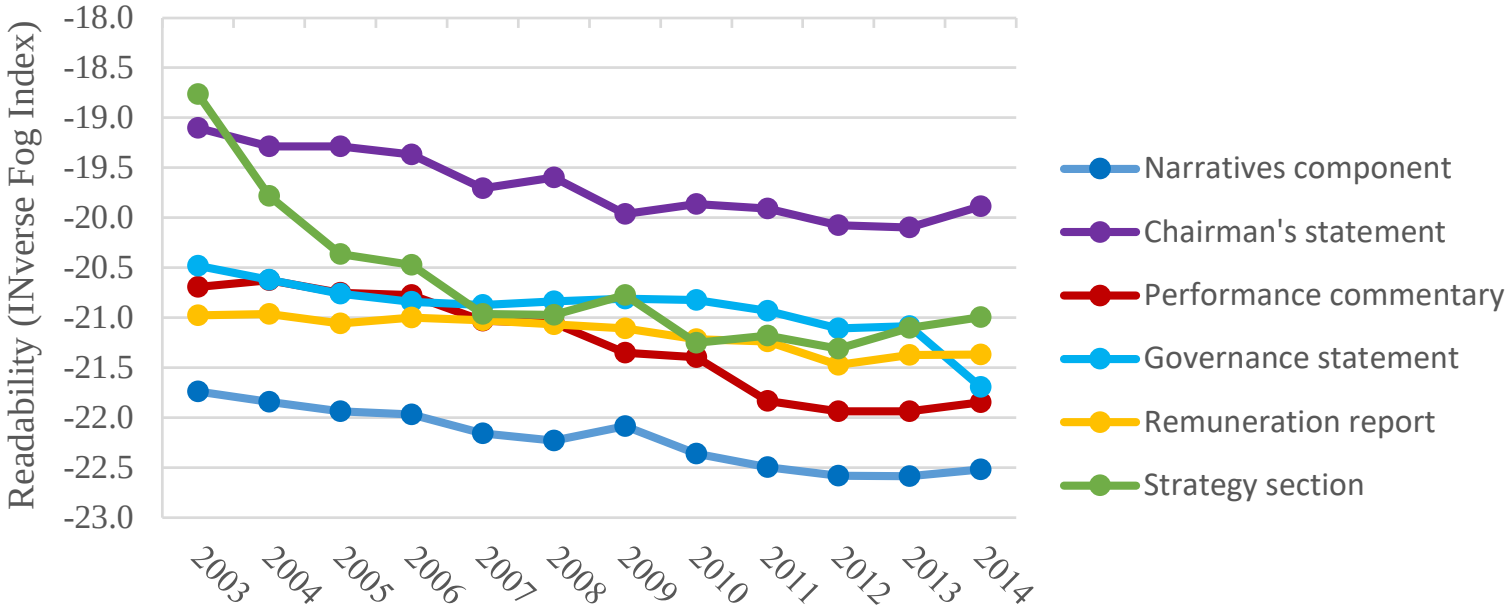
			Error frequencies		Retrieval performance (%)		
	N actual	N classified	Type 1	Type 2	Precision	Recall	F_1 score
Narratives component	4,929	4,846	88	83	98.22	98.32	98.27
Financials component	5,434	5,536	83	88	98.47	98.38	98.43
By section category:							
Chairman's letter	521	517	3	4	99.43	99.23	99.33
CEO review	280	273	10	7	96.55	97.50	97.02
CFO review	328	309	12	19	96.47	94.21	95.33
Governance statement	491	477	27	14	94.79	97.15	95.95
Remuneration report	406	397	0	9	100.00	97.78	98.88
Highlights	276	275	3	1	98.92	99.64	99.28

Reporting growing over Time

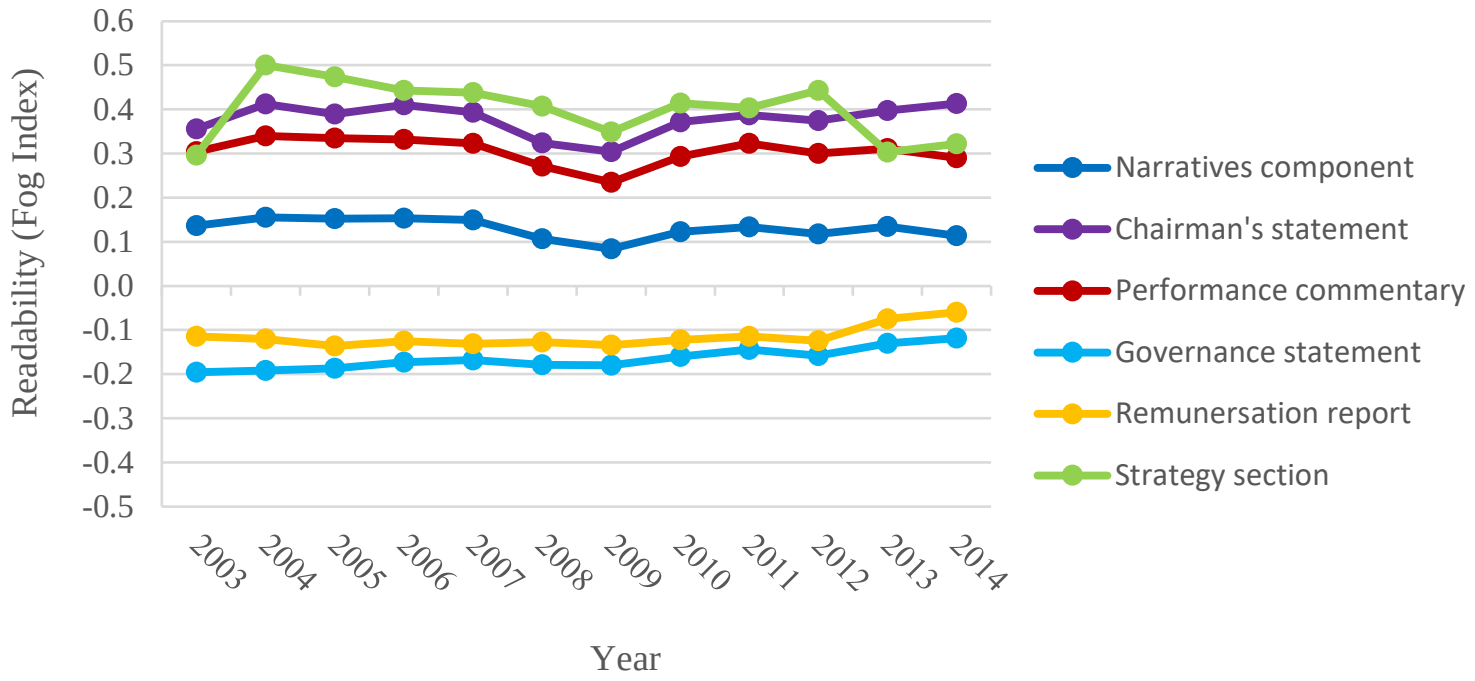
how some of the main types of content have grown over the sample period



Readability (inverse Fog)

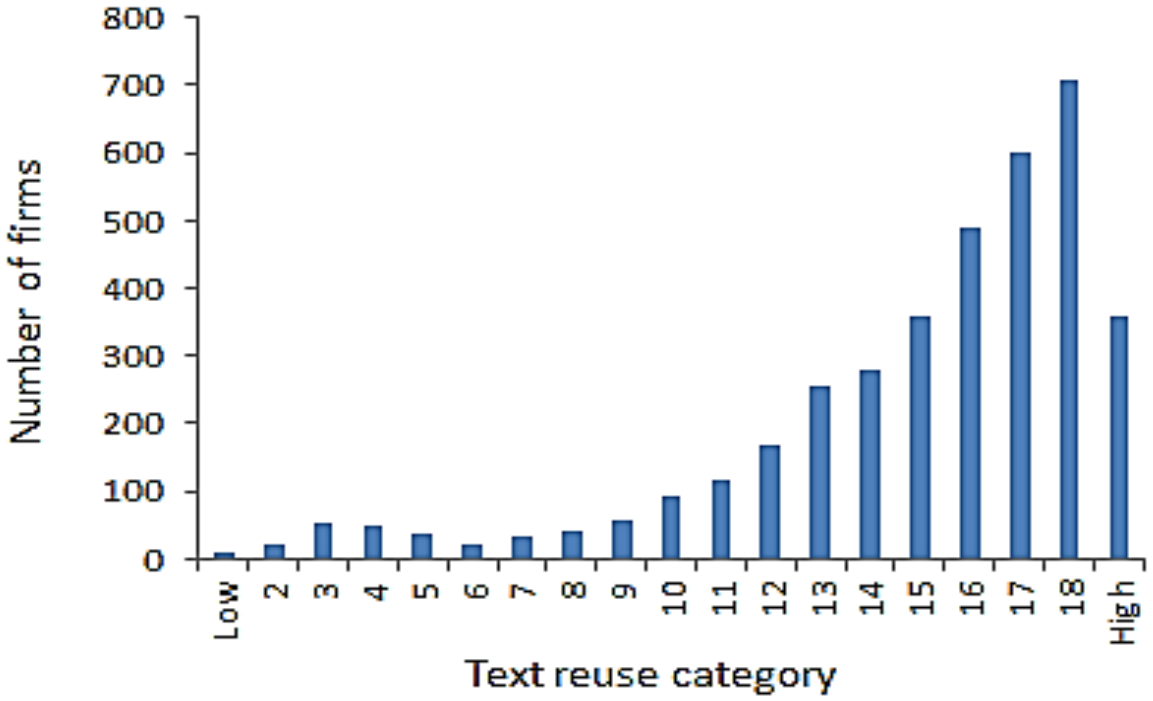


Net Tone


Readability (Fog Index)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year
Narratives component
Chairman's statement
Performance commentary
Governance statement
Remunersation report
Strategy section

Text Reuse (Boilerplating)

Governance statements: Distribution of similarity



Limitations

- Designed for large sample analysis
  - Accuracy of extraction process estimated to be around 90%
  - Extraction process can result in errors at the individual firm level (which are removed or diluted in large sample work)
  - Cannot capture the richness of *how* information is presented (e.g., graphics, charts, tables, etc.)
  - No attempt (yet) to parse text in the financial statements

Summary & Conclusions

- Develop and validate a method for extracting the content and structure of UK annual reports published as PDF files
  - Retaining the structure of the report creates opportunities for research examining new features of disclosure

- Provide the first large-sample evidence regarding the predictive ability of UK annual report narratives
  - Narratives in their entirety are incrementally predictive for earnings beyond financial statement data
  - Difference annual report sections are associated with different predictive qualities
  - Abnormal managerial optimism is associated with lower predictive ability and independent chairman commentaries help to negate managerial bias

Collaboration / Future Work

FRC	•Funding •Help understand UK annual reports
Financial Firm, London	•Fraud Detection •Management Change. Management vs Analysts
Financial Services Provider, London	•Interest in the tool •Automatic Structuring, Indexing Annual Reports
Big Four in Accounting	•Interest in the tool •Provided free-to-publish annual reports
Big Data company, London	•Interest in working together
KSU, Saudi Arabia	•Arabic Financial Narratives (Book Chapter)
MultiLingual Analysis	• Portuguese and Spanish (Book Chapter)
Sunway University, Malaysia	•Malay Financial Narratives
National Natural Science Foundation of China (NSFC)	•Chinese Financial Narratives

Thanks

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

CFIE-FRSE - WEB: https://cfie.lancaster.ac.uk:8443/

CFIE-FRSE – Desktop: https://drelhaj.github.io/CFIE-FRSE

Tutorial

- CFIE-FRSE - WEB:

https://cfie.lancaster.ac.uk:8443/

- Sample Annual Reports:

http://bit.ly/2n2sqcY

Or

http://bit.ly/2nfza8u

- Wmatrix Tutorial:

http://ucrel.lancs.ac.uk/wmatrix/tutorial/