

*PhD Training Session at
Bangor Business School:
Analysing Annual Report
Narratives*

2 December 2014, Bangor

Steven Young
Mahmoud E-Haj

Research funded by ESRC and ICAEW

**Computer-based
Analysis of UK
Annual Report
Narratives**

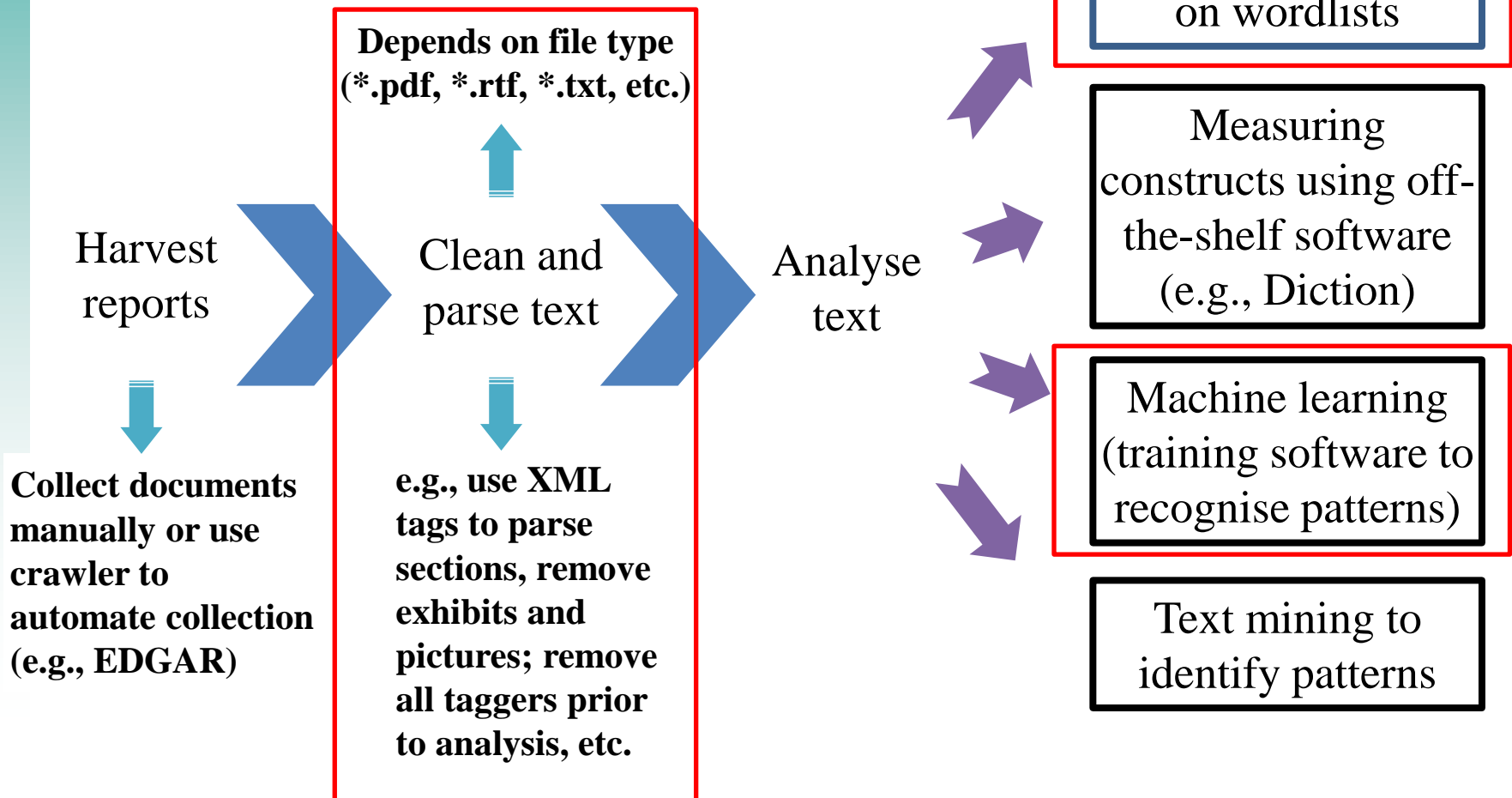
ANALYSING UK ANNUAL REPORT NARRATIVES

Background & objectives

- Part of an ESRC- and ICAEW-funded project examining the *Corporate Financial Information Environment*
 - Martin Walker, *Manchester Business School*
 - Steven Young, *Lancaster University Management School*
 - Paul Rayson, *Lancaster University School of Computing & Communications*
 - Mahmoud E-Haj, *Lancaster University School of Computing & Communications*
 - Vasiliki Athanasakou, *London School of Economics*
- Project seeks to analyse UK financial narratives, their association with financial statement information, and their informativeness for investors
- Automated, large sample analysis of UK annual report narratives represents a cornerstone of the project
 - Develop software for general use by academics

ANALYSING UK ANNUAL REPORT NARRATIVES

Analysing annual report narratives



ANALYSING UK ANNUAL REPORT NARRATIVES

Extant research

- Majority of large sample analysis of annual report narratives has been conducted on US filings (10-Ks) available via EDGAR
 - Management Discussion and Analysis (MD&A) section (Item 7)
 - Risk-related disclosures (Item 1A and Item 7A)
 - Entire 10-K filing
- Analysis of 10-K filings in EDGAR is relatively straightforward
 - Plain text files with consistent structure
 - Use HTML/XML taggers to identify section(s) and extract text
- UK annual reports pose more significant challenges to researchers
 - Normally supplied as *.pdf
 - Unstructured format → no consistent template

ANALYSING UK ANNUAL REPORT NARRATIVES

EDGAR 10-K format extract

PART I

- ITEM 1 Description of Business
- ITEM 1A. Risk Factor
- ITEM 1B. Unresolved Staff Comments
- ITEM 2. Description of Properties
- ITEM 3. Legal Proceedings
- ITEM 4. Mine Safety Disclosures

PART II

- ITEM 5. Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities
- ITEM 6. Selected Financial Data
- ITEM 7. Management's Discussion and Analysis of Financial Condition and Results of Operations
- ITEM 7A. Quantitative and Qualitative Disclosures About Market Risk
- ITEM 8. Financial Statements and Supplementary Data
- ITEM 9. Changes in and Disagreements With Accountants on Accounting and Financial Disclosure
- ITEM 9A. Controls and Procedures
- ITEM 9B. Other Information

ANALYSING UK ANNUAL REPORT NARRATIVES

EDGAR 10-K example

PART III

ITEM 10. Directors, Executive Officers and Corporate Governance

ITEM 11. Executive Compensation

ITEM 12. Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters

ITEM 13. Certain Relationships and Related Transactions, and Director Independence

ITEM 14. Principal Accounting Fees and Services

PART IV

ITEM 15. Exhibits, Financial Statement Schedules Signatures

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual report tool: Extraction

- Use contents page to extract text by section from digital pdf
- Steps in extraction process:
 - Detect contents page
 - Parse contents page
 - Detect page numbering to determine section start/end
 - Add headers as bookmarks to pdf
 - Extract text for each section
- Analyse extracted text by section and for entire document



Contents

Chairman's Statement	02
Chief Executive's Statement	05
Finance Director's Review	08
Directors	12
Directors' Report	13
Corporate Governance	15
Independent Auditors' Report	20
Consolidated Income Statement	22
Consolidated Balance Sheet	23
Company Balance Sheet	24
Consolidated Statement of Recognised Income and Expense	25
Consolidated Reconciliation of Movements in Equity	25
Company Statement of Recognised Income and Expense	25
Company Reconciliation of Movements in Equity	25
Consolidated Cash Flow Statement	26
Company Cash Flow Statement	26
Notes to the Financial Statements	27
Board Report on Directors' Remuneration	54
Notice of Annual General Meeting	58
Shareholder Information	59
Group Five Year Record	60
Corporate Information	Inside Back Cover

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual reports: Unstructured format

Contents		
IFC Financial and operating highlights	56 Consolidated income statement	101 Company cash flow statement
02 Chairman's statement	57 Consolidated balance sheet	Company statement of recognised income and expense
04 Chief executive's report	58 Consolidated cash flow statement	Company statement of changes in equity
08 Performance review	59 Reconciliation of net cash flow to movements in net debt	102 Notes to the Company financial statements
26 Corporate social responsibility report	Consolidated statement of recognised income and expense	112 Five year record
32 Board of directors and company secretary	Consolidated statement of changes in equity	113 Shareholder analysis
33 Directors' report	60 Notes to the consolidated financial statements	114 Financial calendar Company information
36 Corporate governance report	99 Independent auditors' report – Company	115 Investor information
42 Directors' remuneration report	100 Company balance sheet	116 Principal operations
54 Statement of directors' responsibilities		
55 Independent auditors' report – Group		

Financial highlights:

- > Continuing profit before taxation* up 25.3% at £50.1m (2006/07: £40.0m)
- > Profit for the period £34.5m (2006/07: loss £22.5m)
- > Adjusted EPS¹ 20.9% ahead at 7.88p (2006/07: 6.52p)
- > Strong cash management continues to drive low net debt² at £200.2m (2006/07: £174.2m)
- > Robust balance sheet – continuation of share buy-back programme in 2008/09
- > Proposed dividend up 5.9% at 4.50p (2006/07: 4.25p)

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual reports: Unstructured format (cont'd)

What's in our report

Features



Sir Ian Gibson
We are committed to making food shopping fresh, friendly and affordable.
Page 2



Dalton Philips
A clear strategy is in place that is delivering our objectives.
Page 4



Richard Pennycook
Our strong financial performance positions us well for sustainable long term growth.
Page 8



Johanna Waterous
A strong performance culture, long term shareholder value and competitive positioning remain key principles.
Page 46

Also see...

Corporate responsibility review 2011/12



Food with thought

Annual review 2011/12



Making great food affordable

Investor relations website
www.morrisons.co.uk/corporate



Directors' report and business review

Introduction

2 Chairman's statement

Business and strategy review

4 Chief Executive's business and strategy review
8 Group Finance Director's financial review
12 Our strategic objectives

Performance review

26 Key performance indicators
28 Risks and uncertainties
30 Corporate responsibility
33 Our people

Governance

36 Board of Directors and Management Board
40 Corporate governance report
46 Directors' remuneration report
56 General information
59 Statement of Directors' responsibilities

Financial statements

60 Group financial statements
60 Independent auditor's report
61 Consolidated statement of comprehensive income
62 Consolidated balance sheet
63 Consolidated cash flow statement
64 Consolidated statement of changes in equity
65 Group accounting policies
70 Notes to the Group Financial statements

95 Company financial statements
95 Company balance sheet
96 Company accounting policies
99 Notes to the Company financial statements

Investor information

108 Five year summary of results
109 Supplementary information
110 Investor relations and financial calendar

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual report tool: NLP

- In addition to performing text extraction, the tool provides a range of text analysis options:
 - Readability metrics
 - Word counts using pre-determined lists (e.g., forward looking, uncertainty, tone, etc.)
 - Word counts based on user-defined wordlists
 - Comparison with reference corpus (word level and semantic level)
 - Concordance and collocates
 - Upload and analyse user-defined text file
- Demo to illustrate functionality →

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual report tool: Demo overview

- Overview (wordlists, readability metrics) and interface with WMatrix
- Uploading one or more annual reports and generating output
- Uploading and analysing with a user-defined key word list
- Uploading and analysing a user-defined text file
- Examples of further analysis in MMatrix:
 - Cloud for chairman's statement vs. standard reference corpus: word level
 - Cloud for chairman's statement vs. to standard reference corpus: semantic level
 - Cloud for chairman's statement vs. chairman's statement corpus: word level
 - Cloud for chairman's statement vs. chairman's statement corpus: semantic level
 - Concordance/collocation

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual report tool: Developments

- The tool also offers the following features:
 - Choice of different readability metrics
 - Text re-use metric to detect boilerplating and incremental changes
 - Method for splitting “front end” and “back end” of the annual report
 - Method for isolating performance commentary (e.g., CEO Review/Business Review/Operating Review/Financial Review)
 - Reference corpora for:
 - **Full Annual Report and Accounts**
 - **Chairman’s Statement**
 - **CEO Review/Business Review/Operating Review/Financial Review**
 - **Corporate Governance**
 - **Directors’ Remuneration**
 - Readability scores for all UK annual reports for the period 2003-2014
 - Method for linking scores to financial and market data from Datastream

ANALYSING UK ANNUAL REPORT NARRATIVES

UK Annual report tool

Questions?

FURTHER DETAILS ON METHODOLOGY

Appendix: Identifying and parsing the contents page

- Identifying and extracting the AR contents page creates a number of challenges
 - Location and format of the contents table is entirely at the discretionary of management → rarely positioned at the beginning of the document (i.e. p.1)
Solution: use a matching algorithm based on set of words and phrases common to tables of contents

FURTHER DETAILS ON METHODOLOGY

Appendix: Identifying and parsing the contents page

- Identifying and extracting the AR contents page creates a number of challenges
 - Location and format of the contents table is entirely at the discretionary of management → rarely positioned at the beginning of the document (i.e. p.1)
Solution: use a matching algorithm based on set of words and phrases common to tables of contents
 - Widespread practice of presenting other text such as company overview and performance highlights on the same page as the table of contents
Solution: extract lines of text from the contents page that start or end with a number between one and the number of pages in the AR

FURTHER DETAILS ON METHODOLOGY

Appendix: Identifying and parsing the contents page

- Identifying and extracting the AR contents page creates a number of challenges
 - Location and format of the contents table is entirely at the discretionary of management → rarely positioned at the beginning of the document (i.e. p.1)
Solution: use a matching algorithm based on set of words and phrases common to tables of contents
 - Widespread practice of presenting other text such as company overview and performance highlights on the same page as the table of contents
Solution: extract lines of text from the contents page that start or end with a number between one and the number of pages in the AR
 - Page numbers referenced in the published table of contents rarely align with page numbers in the PDF
Solution: use an algorithm that crawls through a dynamic set of 3 pages to identify a pattern of sequential numbers with increment one (e.g. 31, 32, 33)

FURTHER DETAILS ON METHODOLOGY

Appendix: Booklet-style annual reports

- Some ARs are published with two report pages (portrait orientation) presented side-by-side on a single page (landscape orientation) in the PDF
 - Booklet-style ARs compound the problem of page number misalignment because two pages in the annual report equate to a single page in the PDF

Solution: unable to devise a reliable algorithm to resolve this pagination problem

FURTHER DETAILS ON METHODOLOGY

Appendix: Booklet-style annual reports

- Some ARs are published with two report pages (portrait orientation) presented side-by-side on a single page (landscape orientation) in the PDF
 - Booklet-style ARs compound the problem of page number misalignment because two pages in the annual report equate to a single page in the PDF

Solution: unable to devise a reliable algorithm to resolve this pagination problem
- Absent a solution for resolving this problem we create a flag to identify possible booklet-style (ARs) → allows user to exclude suspect reports
- Use two conditions to classify annual reports as booklet-style candidates:
 - Cases where a contents header including the keywords “notes to” or “notes relating to” corresponds to a page number > number of pages in the PDF file
 - Page width in PDF file > 800 points

FURTHER DETAILS ON METHODOLOGY

Appendix: Treatment of image-based PDFs

- Most U.K. firms now provide their AR as a digital PDF file
 - But prior to 2006, a non-trivial fraction of ARs were supplied as image-based (i.e., scanned) PDFs
 - Text extraction methods cannot be applied to image-based PDF files
- Sought to remedy this problem by converting all image-based files to digital format using optical character recognition (OCR) techniques
- Problem: tests revealed that OCR does not convert text in a way that permits tables of contents to be parsed reliably
 - While converted tables of contents appear to be structured appropriately, actual flow of text often reads down a column rather than across each row
- Inability to process image-based PDFs biases against smaller firms before 2006

FURTHER DETAILS ON METHODOLOGY

Appendix: Distinguishing narratives from financial statements

- Primary focus relates to the narrative component of the AR
 - Isolating the narrative element is a non-trivial task due to the lack of standardization with respect to the ordering and labelling of section headers
- Use a two-step approach that involves:
 - A preliminary split based on the structure of a representative U.K. annual report;
 - Followed by an updating process that accounts for deviations from our representative case

FURTHER DETAILS ON METHODOLOGY

Appendix: Distinguishing narratives from financial statements

1. Section	Comments
2. Overview	(including highlights)
3. Chairman's statement	
4. Performance commentary	(including one or more of the following sections: chief executive's review, review of operations, business review, strategic review, financial review)
5. Other sections	(various, common examples of which include risk review and corporate social responsibility report)
6. Director's biographies	
7. Directors' report	
8. Governance statement	
9. Remuneration Report	
10. Statement of directors' responsibilities	
11. Auditor's report	
12. Primary financial statements	(as required by IAS 1)
13. Notes to the accounts	(as required by IAS 1)
14. Other disclosures	(various, common examples of which include notice of annual general meeting, three- or five-year review, subsidiaries and operating locations, etc.)

FURTHER DETAILS ON METHODOLOGY

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



FURTHER DETAILS ON METHODOLOGY

Appendix: Isolating performance commentary

- Performance commentary is a particularly important category of the narrative section of U.K. ARs
 - Equivalent to Item 7: Management Discussion and Analysis (MD&A) in 10-K
- Unlike 10-K filings, structure and format of performance reporting in U.K. ARs is largely discretionary
 - Some firms limit performance commentary to a statement by the CEO reviewing operational matters and a report by the CFO on financial
 - Majority of firms decompose performance-related discussions across multiple sections with non-standard headings
- Use an algorithm that harvests all sections between pre-specified start (s) and end (e) points based on the representative AR structure
 - Supplementary adjustments to deal with reports that deviate from the base case

FURTHER DETAILS ON METHODOLOGY

Appendix: Isolating performance commentary

1. Section	Comments
2. Overview	(including highlights)
<i>S</i> → 3. Chairman's statement	
4. Performance commentary	(including one or more of the following sections: chief executive's review, review of operations, business review, strategic review, financial review)
5. Other sections	(various, common examples of which include risk review and corporate social responsibility report)
<i>e</i> { 6. Director's biographies	
7. Directors' report	
8. Governance statement	
9. Remuneration Report	
10. Statement of directors' responsibilities	
11. Auditor's report	
12. Primary financial statements	(as required by IAS 1)
13. Notes to the accounts	(as required by IAS 1)
14. Other disclosures	(various, common examples of which include notice of annual general meeting, three- or five-year review, subsidiaries and operating locations, etc.)

FURTHER DETAILS ON METHODOLOGY

Example

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

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

s →

e →

←

ANALYSIS OF STRATEGY-RELATED CONTENT

Summary

- Automated large sample analysis of non-US annual reports is hampered by format and structure
- Develop a free-to-access software tool for extracting and analysing UK annual report narratives (plus text from other sources)
 - Extracts from pdfs by section
 - Provides simple NLP metrics such as readability, word counts, etc.
 - Interfaces with MWatrix for more sophisticated linguistic analysis at parts of speech and semantic levels
- Functionality of software expected to develop significantly over coming months
- Using methods to examine a series of disclosure-related questions