

Corporate Financial Information Environment (CFIE)



Prof Martin Walker
Dr Mahmoud El-Haj

Prof Stephen Young
Dr Vasiliki Athanasakou

Dr Paul Rayson
Dr Thomas Schleicher

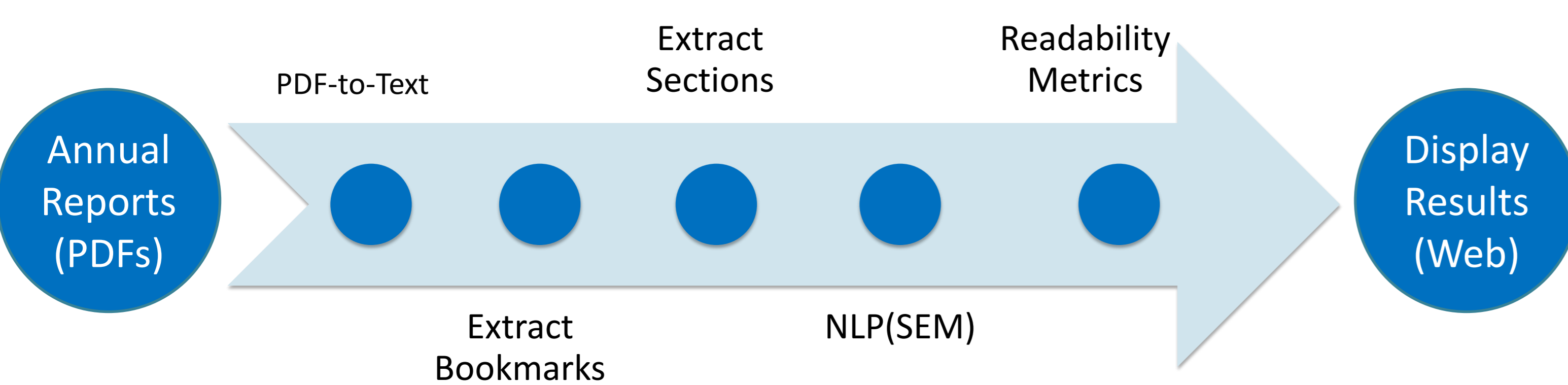
PURPOSE AND MOTIVATION

Study the causes and consequences of corporate disclosure and financial reporting outcomes. The determinants of financial reporting quality and the factors that influence the quality of information disclosed to investors beyond the financial statements are the main focal points of the proposed project.

DATA COLLECTION

- Annual Reports of UK Firms.
- 250 Searchable PDFs.
- 50 UK Firms.
- 5 annual reports over 5 years.
- 2005-2012.

NLP AND READABILITY TOOLS



EXTRACTION TOOLS AND MEASUREMENTS

- **Open Source**
 - XPDF
 - PDF-to-HTML
 - LA-PDF-Text
 - Apache PDFBox
 - Improved Apache PDFBox
 - Multivalent
 - iText ✓
 - PDFTextStream ✓
- **Commercial**
 - Adobe X Pro ✓
 - OmniPage
- **Readability (Phantom, Java)**
 - Flesch-Kincaid
 - Gunning Fog Index
- **Other Dimensions (word-lists)**
 - Forward Looking
 - Hedges
 - Tone (Positive, Negative)
 - Uncertainty
 - Challenges

EXTRACTED SECTIONS

- **Annual Report**
 1. Chairman's statement
 2. Directors Report
 3. CEO Review
 4. Corporate Government Report
 5. Directors Remuneration Report
 6. Directors Report and Business Review
 7. Directors Responsibilities Statement
 8. Financial Review
 9. Key Performance Indicator
 10. Operating Review
 11. Highlights
- **Financial Statement**
 1. Independent auditors' report.
 2. Mandatory financial statements (including statement of financial position, cash flow statement)
 3. Other information (including information for shareholders...etc.)

PREPROCESSING

To ease the matching process, the headers had to undergo the following steps (using Regular Expressions):

1. Convert all headers to lowercase.
2. Remove punctuations (e.g. single-quote).
3. Get the singular or the plural form of headers text (e.g. "director", "chairman", "responsibility" - instead of "directors", "chairman's", "responsibilities").

LEVENSHTEIN DISTANCE

We modified the algorithm to work on a word level instead of character level.

This transforms the algorithm's meaning to: the minimum number of word edits (insertion, deletion, substitution) required to change one sentence into the other.

EXTRACTION PROCESS

Example: "Chairman's statement"

1. Select all headers containing any of the chairman header's words (e.g. "Chair", "statement", "introduction", "letter to"...etc).
2. Provide the list to Accounting and Finance expert to create gold-standards (synonyms list).
3. Insert the headers & their info into a database-table.
4. Run the selected headers through a decision-making (keep one chairman's header per firm).

CORPORATE GOVERNMENT REPORT (SAMPLE)

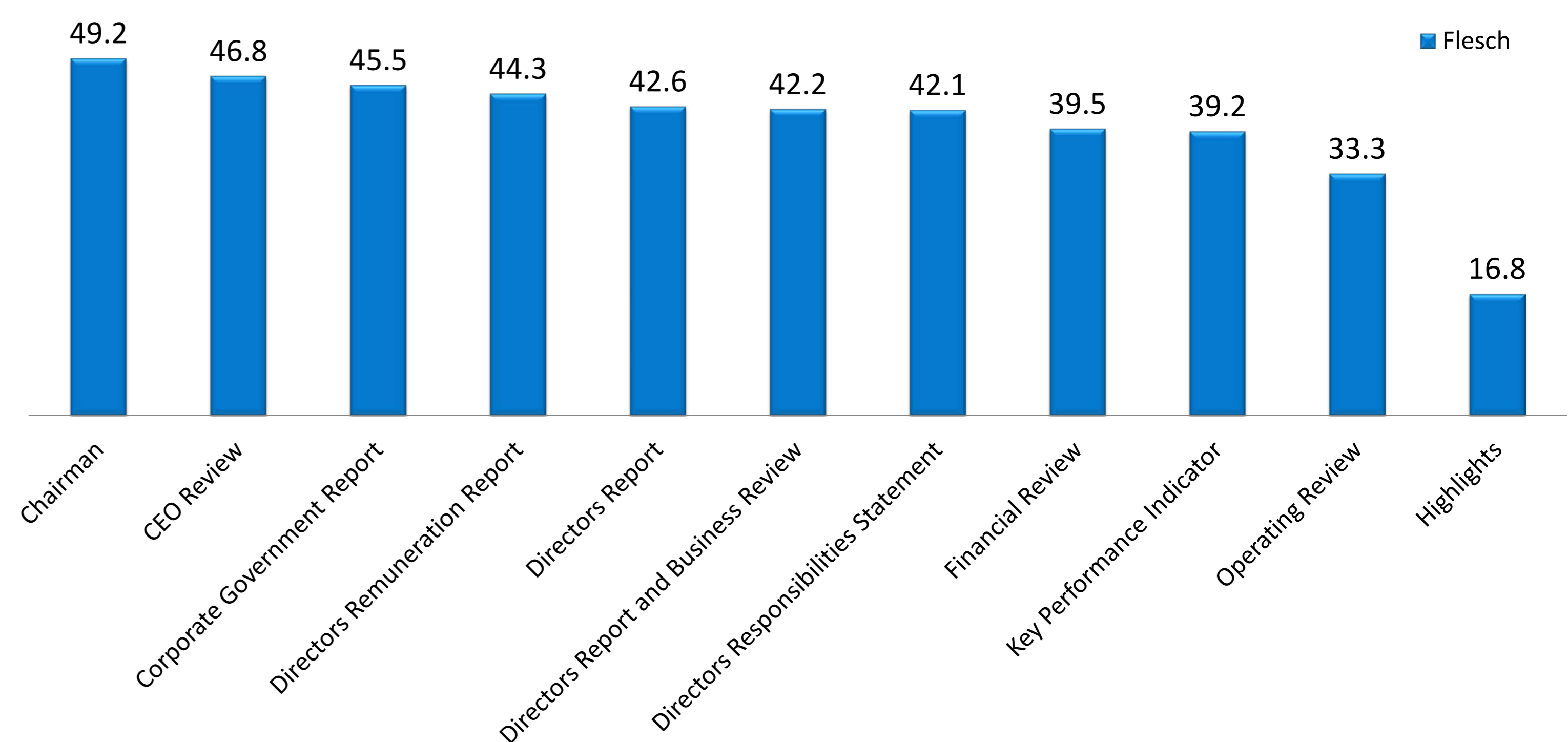
- List of Synonyms:
 - corporate governance report
 - corporate governance statement
 - corporate governance
 - director statement on corporate governance
 - corporate governance review
 - report on corporate governance
 - director report on corporate governance
 - statement on corporate governance
- Total Matches: 232
- Exact Matches: 232
- Partial Matches: 0

ERROR ANALYSIS

Some annual reports:

- do not contain bookmarks.
- bookmarks available are incorrect/non-textual.
- do not contain any of the headers and their synonyms.

SECTIONS READABILITY MEASURE



ANNUAL REPORT VS. CHAIRMAN'S STATEMENT

