Textual Analysis of Earnings Calls: Differences between the Firms

Paper Number: *2022-EU-30MP-1056*

Authors:
Nilofar Varzgani (Presenter)
Assistant Professor of Business Systems and Analytics
La Salle University, Philadelphia, PA USA

Yusuf Joseph Ugras
Interim Dean and Associate Professor of Accounting, La Salle University

Özgür Arslan Ayaydin
Clinical Professor Department of Finance, University of Illinois Chicago

Introduction and Motivation

- Accounting numbers influence decision making by investors.
- Qualitative analytics from the firms and from the media influences investor behavior.
  - Texts in 10-K Reports,
  - Earnings Press Releases,
  - **Earnings Conference Calls**,
  - Comment Letters to the SEC,
  - Analysts Remarks,
  - Articles in Media and
  - Conversations in Social Media

Textual Analytics in Academic Literature

- **Fog Index –** average sentence length and word complexity
- **Length of Reports –** file size
- **Word Lists -** measure tone optimism/pessimism
  - Henry Word List (2008),
  - Harvard General Inquirer Word List (GI),
  - Diction Optimism and Pessimism Word Lists, and
  - Loughran and McDonald Word List (2011)
    - » 354 positive and 2,329 negative words
    - » created with financial communication in mind
    - » Used by several studies to analyze financial texts

Data

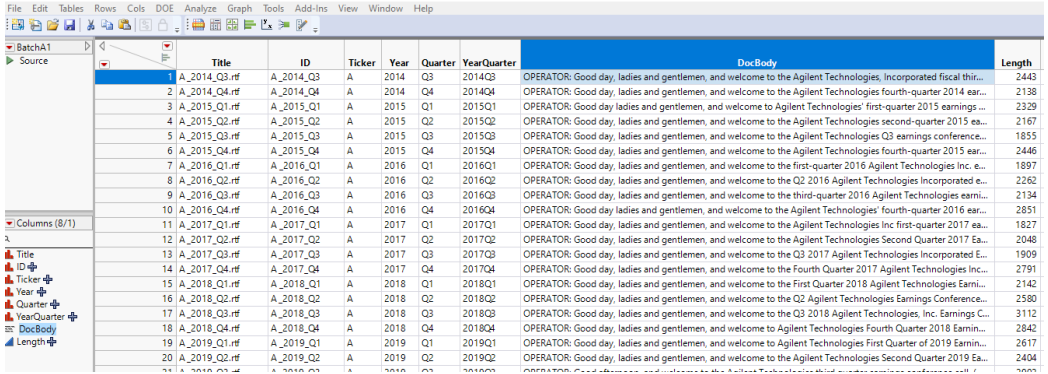
- My data sample consists of U.S. data of SNP 1500 publicly listed firms for the period of **2007Q1** up to **2020Q4**.
- All earnings call transcripts were downloaded from LexisNexis in ".rtf" format.
- These transcripts are void of any graphics and special characters.
- After data cleaning and preprocessing, I ended up with a sample of 24,903 earnings call transcripts spanning over 2007-Q1 to 2020-Q4.
- Earnings conference calls usually have two main segments.
- **Structure of the Earning Conference Call:**

1. First Line: Title
2. Second Line: "FD (Fair Disclosure) Wire"
3. Third Line: Date in the format "May 6, 2009 Wednesday"

- Participants Section
- Presentation Section
- Questions and Answers Section

Methodology

- Extracted the transcripts' prepared remarks section as DocBody, QnA as the Discussion, and number of analysts attending as the analyst coverage for each of the files.
- Created columns for ID, Ticker, Year, Quarter, YearQuarter, and Length (count of number of words in DocBody). Distribution of Length is given the output below.
- Changed the column ***data type*** for DocBody to Character with Modeling Type as Unstructured Text for the Text Explorer Platform to work.

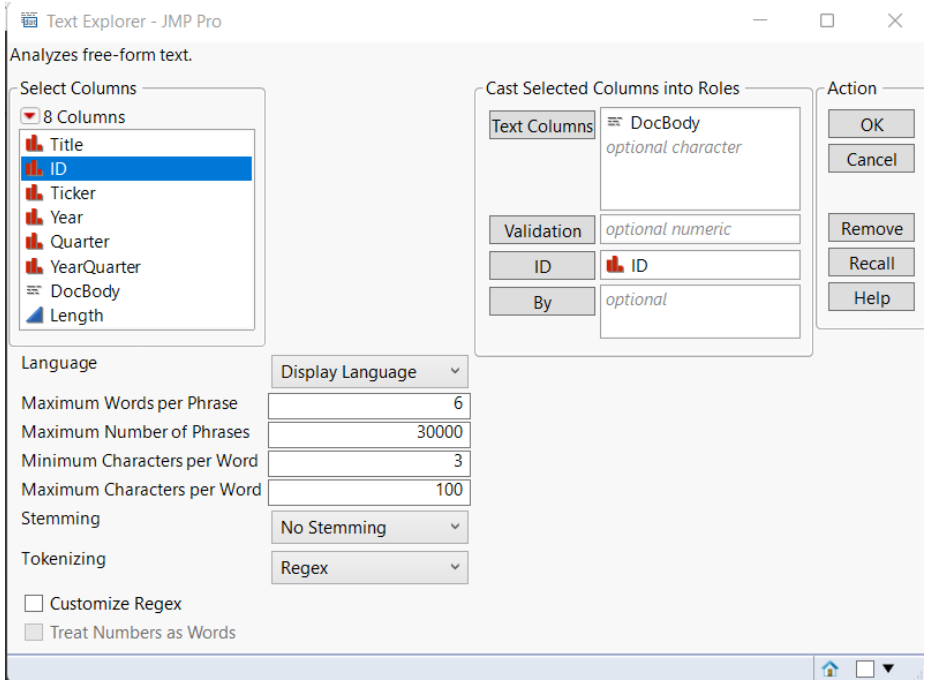


	Title	ID	Ticker	Year	Quarter	YearQuarter	DocBody	Length
1	A_2014_Q3.rtf	A_2014_Q3	A	2014	Q3	2014Q3	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies, Incorporated fiscal thir...	2443
2	A_2014_Q4.rtf	A_2014_Q4	A	2014	Q4	2014Q4	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies fourth-quarter 2014 ear...	2138
3	A_2015_Q1.rtf	A_2015_Q1	A	2015	Q1	2015Q1	OPERATOR: Good day ladies and gentlemen, and welcome to Agilent Technologies' first-quarter 2015 earnings ...	2329
4	A_2015_Q2.rtf	A_2015_Q2	A	2015	Q2	2015Q2	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies second-quarter 2015 ea...	2167
5	A_2015_Q3.rtf	A_2015_Q3	A	2015	Q3	2015Q3	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies Q3 earnings conference...	1855
6	A_2015_Q4.rtf	A_2015_Q4	A	2015	Q4	2015Q4	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies fourth-quarter 2015 ear...	2446
7	A_2016_Q1.rtf	A_2016_Q1	A	2016	Q1	2016Q1	OPERATOR: Good day, ladies and gentlemen, and welcome to the first-quarter 2016 Agilent Technologies Inc. e...	1897
8	A_2016_Q2.rtf	A_2016_Q2	A	2016	Q2	2016Q2	OPERATOR: Good day, ladies and gentlemen, and welcome to the Q2 2016 Agilent Technologies Incorporated e...	2262
9	A_2016_Q3.rtf	A_2016_Q3	A	2016	Q3	2016Q3	OPERATOR: Good day, ladies and gentlemen, and welcome to the third-quarter 2016 Agilent Technologies earni...	2134
10	A_2016_Q4.rtf	A_2016_Q4	A	2016	Q4	2016Q4	OPERATOR: Good day ladies and gentlemen, and welcome to the Agilent Technologies' fourth-quarter 2016 ear...	2851
11	A_2017_Q1.rtf	A_2017_Q1	A	2017	Q1	2017Q1	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies Inc first-quarter 2017 ea...	1827
12	A_2017_Q2.rtf	A_2017_Q2	A	2017	Q2	2017Q2	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies Second Quarter 2017 Ea...	2048
13	A_2017_Q3.rtf	A_2017_Q3	A	2017	Q3	2017Q3	OPERATOR: Good day, ladies and gentlemen, and welcome to the Q3 2017 Agilent Technologies Incorporated E...	1909
14	A_2017_Q4.rtf	A_2017_Q4	A	2017	Q4	2017Q4	OPERATOR: Good day, ladies and gentlemen, and welcome to the Fourth Quarter 2017 Agilent Technologies Inc...	2791
15	A_2018_Q1.rtf	A_2018_Q1	A	2018	Q1	2018Q1	OPERATOR: Good day, ladies and gentlemen, and welcome to the First Quarter 2018 Agilent Technologies Earni...	2142
16	A_2018_Q2.rtf	A_2018_Q2	A	2018	Q2	2018Q2	OPERATOR: Good day, ladies and gentlemen, and welcome to the Q2 Agilent Technologies Earnings Conference...	2580
17	A_2018_Q3.rtf	A_2018_Q3	A	2018	Q3	2018Q3	OPERATOR: Good day, ladies and gentlemen, and welcome to the Q3 2018 Agilent Technologies, Inc. Earnings C...	3112
18	A_2018_Q4.rtf	A_2018_Q4	A	2018	Q4	2018Q4	OPERATOR: Good day, ladies and gentlemen, and welcome to Agilent Technologies Fourth Quarter 2018 Earnin...	2842
19	A_2019_Q1.rtf	A_2019_Q1	A	2019	Q1	2019Q1	OPERATOR: Good day, ladies and gentlemen, and welcome to Agilent Technologies First Quarter of 2019 Earnin...	2617
20	A_2019_Q2.rtf	A_2019_Q2	A	2019	Q2	2019Q2	OPERATOR: Good day, ladies and gentlemen, and welcome to the Agilent Technologies Second Quarter 2019 Ea...	2404

Textual Analytics
JMP Terminology

- A **term** or **token** is the smallest piece of text, similar to a word in a sentence. The process of breaking the text into terms is called tokenization.
- A **phrase** is a short collection of terms; the platform has options to manage phrases that are specified as terms in and of themselves.
- A **document** refers to a collection of words; in a JMP data table, the unstructured text in each row of the text column corresponds to a document.
- A **corpus** refers to a collection of documents.
- It is often desirable to exclude some common words from the analysis. These excluded words are called **stop words**.
- **Stemming** is the process of combining words with identical beginnings (stems) by removing the endings that differ. This results in “jump”, “jumped”, and “jumping” all being treated as the term “jump·”.
  - No Stemming: No terms are combined.
  - Stem for Combining: Stems only the terms where two or more terms stem to the same term.
  - Stem All Term: Stems all terms.

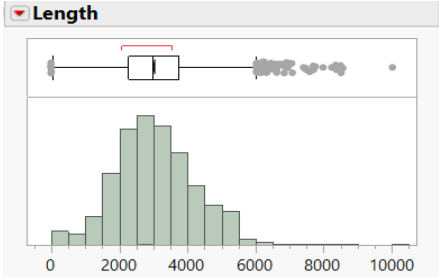
Text Explorer Platform in JMP Pro 16



Initial Output

- Next, I reviewed the list of phrases and selected phrases that should be used as terms.
- There were a total of 30,000 phrases, out of which I added 1,068 phrases to the terms list.
- All phrases used less than 900 times in the full sample were ignored.
- After adding the phrases to the terms, I then reviewed the terms in order to decide which terms should be converted to stop words.
- Import the list of all terms and then filter them out in Excel to
  - first identify numbers being read as text and add those as stop words.
  - identify text with total count less than 10.
  - identify month names and remove all variations of those except for March since the COVID-19 pandemic was announced in March so I left March in. I also left March, marching, and marched in as well due to the verb.
  - Next, I went through the term list again and manually selected stop words one last time before finalizing the word cloud.

Output- Distribution of *Length*

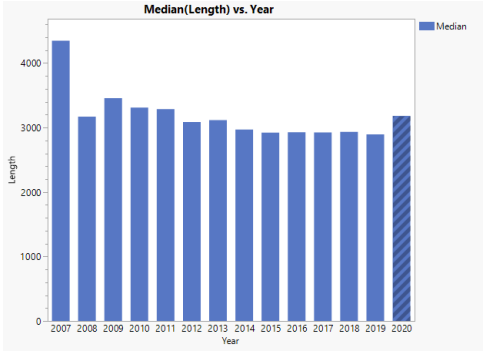


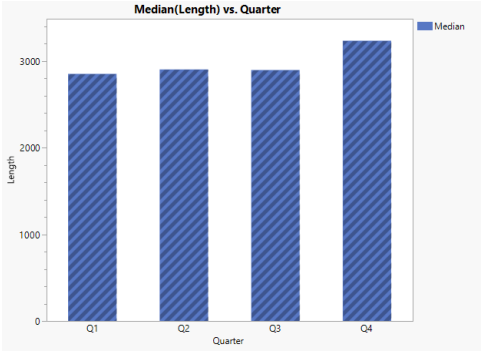
Quantiles

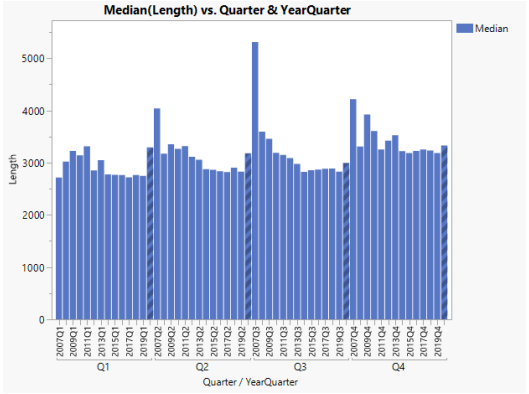
100.0%	maximum	10008
99.5%		5767.32
97.5%		5236
90.0%		4569.6
75.0%	quartile	3764
50.0%	median	2966
25.0%	quartile	2266
10.0%		1687.4
2.5%		790
0.5%		101
0.0%	minimum	1

Summary Statistics

Mean	3027.3539
Std Dev	1125.6666
Std Err Mean	7.1331925
Upper 95% Mean	3041.3353
Lower 95% Mean	3013.3724
N	24903







Top Ten Tickers in terms of Length

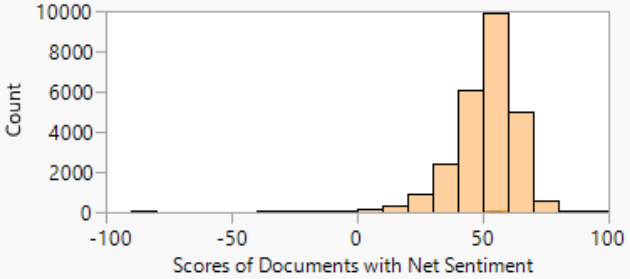
Tickers	Average of Length
BXP	5915
KMP	5681
SIGI	5635
TGT	5225
WSM	5221
BAC	5206
AZO	5187
CXW	5162
NKE	5118
GE	5111

Word Cloud
(With and Without Stemming)

Sentiment Analysis Output

Sentiment Summary

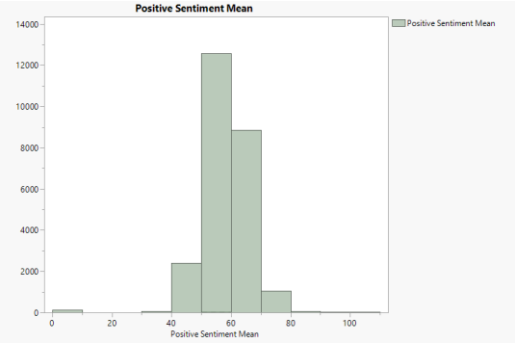
	N	Mean Score
All Scored Documents	24810	33.9
Net Positive Documents	24744	51.4
Net Negative Documents	60	-12.3
No Sentiment Documents	93	0.0

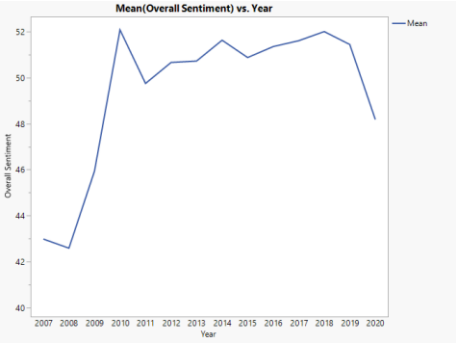


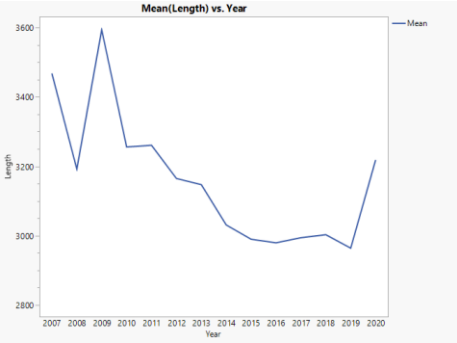
Document	Positive Sum	Positive Score Mean	Negative Sum	Negative Score Mean	Overall Score
4	817	63	0	0	63
5	494	62	0	0	62
6	364	61	0	0	61
7	390	56	0	0	56
8	220	55	0	0	55
9	150	50	0	0	50
10	643	54	-20	-20	48
11	334	56	0	0	56
12	729	73	0	0	73
13	291	58	0	0	58

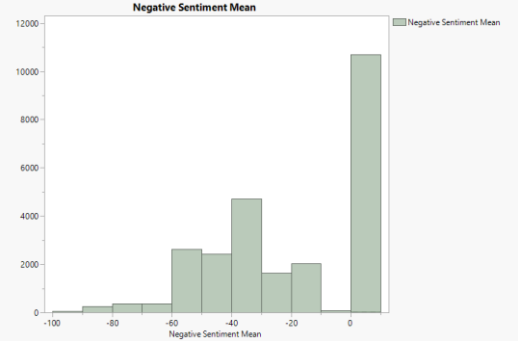
Sentiment	Score	Count
good	60	89209
pleased	40	29787
better	60	29693
great	80	23372
best	90	20663
working	30	20024
favorable	40	18469
success	50	18362
excited	75	11417
successful	50	11175
excellent	90	10651

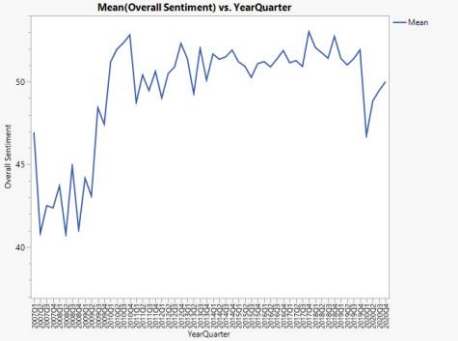
Sentiment Analysis Output

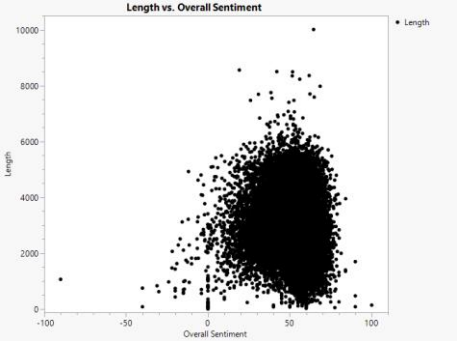












Companies with Highest and Lowest Mean Overall Sentiment

Highest

Company	Industry
Apple	Technology
Ametek	Electronic Instruments
Amphenol Corporation	Electronics and Fiber Optics
Electronic Arts	Gaming
Best Buy	Electronic Retailer
Synopsys	Electronic Design Automation
PACCAR	Automotive (Trucks)
TJX Companies	Departmental Stores
Salesforce	Technology
Walker & Dunlop	Financial Services

Lowest

Company	Industry
Avanos Medical	Medical Technology
Mercury General Corporation	Insurance
Stericycle	Medical Waste Management
Waste Management	Waste Management
O-I Glass	Container Glass Products
PetMed Express	Online Pet Pharmacy
Trimble Inc	Technology
Republic Services	Waste Management
Zynex	Medical Technology
Associated Banc-Corp	Retail Banking

Future Research

- Examine the tone of the earnings calls by cross analysis with:
  - Managerial strategic incentives
  - Disclosures
  - Impact on analysts and investors
  - Firm size, complexity, age etc.
- Explore term selection for building data mining models.