

TALKING ABOUT REGULATION IN 10-K ANNUAL REPORTS;
UNIFORMITY IN A NAIVE SAMPLE.

by

KATERINA RUTH MILLS, ESQ.,

B.A., University of California Los Angeles, 2011

J.D., University of Denver, 2013

M.A., University of Denver, 2014

B.A., University of Denver, 2015

A Thesis submitted to
the Faculty of the Leeds School of Business
University of Colorado, Boulder in partial fulfillment
of the requirements for the Degree
Master of Science
Department of Finance

2019

This thesis entitled:
Talking about Regulation in 10-K Annual Reports; uniformity in a naive sample.
Written by Katerina Ruth Mills
has been approved by the Department of Finance

Professor Edward Van Wesep, Committee Chair

Professor Diego Garcia, Committee Member

Professor Tony Cookson, Committee Member

Date: _____

The final copy of this thesis has been examined by the above signatories, and we find that both the content and form meet acceptable presentation standards of scholarly work in the above-mentioned discipline.

Author: Katerina Ruth Mills (M.S.B.A., Finance, Leeds School of Business, Department of Finance)
Title: Talking about Regulation in 10-K Annual Reports; uniformity in a naive sample.
Committee: Edward D. Van Wesep, PhD (Chair), Diego Garcia, PhD, and J. Anthony Cookson, PhD

Abstract

This project seeks to highlight the difficulties of employing required annual reports in analyses attempting to tie left-hand side outcomes, whether past or future, through the use of Natural Language Processing techniques to analyze firm discussion of regulators, regulation, laws, and other regulatory regimes in the context of required 10-K annual disclosures under U.S. public company reporting regimes governed by the U.S. Securities and Exchange Commission (SEC). While Natural Language Processing (NLP) techniques have gained popularity in turning text into data facilitating a multitude of varied new analytical projects in the fields of academic corporate research, using NLP mechanisms particularly for regulation-oriented corporate speech analysis presents relative uniformity across filers and industries which this small project seeks to highlight as a possible burden for NLP usage in this particular extension of the legal and financial reporting analyses. The small sample, simple word lists, and naive comparison employed here seeks to highlight, through simple methodology, that even an industry-by-industry analysis method may be less than meaningful in addressing regulatory, industry standard, and legal practice nuances defining regulatory discussion due, at least in part, to uniform reporting summaries (relative uniformity) across firms, not only those participating in a single industry, but across SEC annual filers generally. This project will conclude with a brief summary of possible causal factors in the highlighted report uniformities while refraining from any implication that regulatory or legal comparison is or should be among the comparative factors for which design and response uniformity guidelines for 10-K reports are generated or sought by the SEC for use by potential or current investors as an indicative factor of firm performance or future performance.

All you really need to know for the moment is that the universe is a lot more complicated than you might think, even if you start from a position of thinking it's pretty damn complicated in the first place.
Douglas Adams

Acknowledgments

Special and unending thanks to Professors Edward D. Van Wesep, Diego Garcia, and J. Anthony (Tony) Cookson whose patience, guidance, and expertise have been and continue to be invaluable in constructing and finishing this project, through the past two years of my graduate study, and, importantly, in my choice and ability to defend this small thesis. Special personal thanks to Stefanie Ungphakorn, without whom the last two years would not have been surmountable.

Additional thanks go to those who are responsible for maintaining the Securities and Exchange Commission and reporting specific regulatory databases, summaries, and filing practice handbooks employed to finish this project. Without their efforts and provision, navigating the complex system of regulations and regimes over the entities and firms reporting their 10-K filings each year would be all but impossible.

Thanks, no matter how special or unending, could never be enough to truly thank my mother and father for their constant support, guidance, and unending patience. And finally, particular thanks to my wonderful spouse, Michael, whose unconditional support and understanding are constant and without limit.

“You know,” said Arthur, “it’s at times like this, when I’m trapped in a Vogon airlock with a man from Betelgeuse, and about to die of asphyxiation in deep space that I really wish I’d listened to what my mother told me when I was young.”

“Why, what did she tell you?”

“I don’t know, I didn’t listen.”

Douglas Adams

TABLE OF CONTENTS

INTRODUCTION	1
SECTION I	3
<i>Brief Literature Review</i>	
<i>10-K Reporting Requirement and Components</i>	
SECTION II	13
<i>Data Summary</i>	
<i>NLP Program</i>	
SECTION III	16
<i>All Filers Data and Initial Summary Remarks</i>	
SECTION IV	21
<i>Industry Comparison</i>	
SECTION V	26
<i>Concluding Remarks</i>	
BIBLIOGRAPHY	28
APPENDICES	32*
<i>Appendix 1: NLP Code</i>	
<i>Appendix 2: Sample Out File</i>	
<i>Appendix 3: SEC SIC Firm Groups List</i>	
<i>Appendix 4: 3 Group SIC Firm Group Comparison Full Data Table</i>	
<i>*Appendix materials begin page 34</i>	

"I thought," he said, "that if the world was going to end we were meant to lie down or put a paper bag over our head or something."

"If you like, yes," said Ford.

"Will that help?" asked the barman.

"No," said Ford and gave him a friendly smile.

Douglas Adams

LIST OF TABLES AND FIGURES

All Filers Data Table	16
 All Filers Figures	
Total 10-K Reports Filed	17
Average Total Paragraphs per Filings	17
Average Regulatory Paragraphs per Filings	18
Average Regulatory Words Flagged per Filings	18
Average Regulatory Paragraphs Total Length per Filing All years	19
Average Regulatory Paragraphs Total Length per Filing 1998-2009	19
Average Total Number of Blame Words per Filing	20
 SIC Comparison Figures	
Total Filings	22
Average Paragraphs	22
Average Regulatory Paragraphs	23
Average Regulatory Words Flagged	23
Average Regulatory Paragraph String Length	24
Average Blame Words Counted in Regulatory Paragraphs	24

“A learning experience is one of those things that says, 'You know that thing you just did? Don't do that.'”
Douglas Adams

INTRODUCTION

10-K disclosures have been the focus of numerous and varied Natural Language Processing (NLP) endeavors inside finance and accounting literature since NLP tools, as mechanisms for turning words into data, infiltrated the academic corporate research community. This small project seeks to highlight and describe an embedded framework of report uniformity which may require special attention or additional specific controls or disclaimers for the promulgation of NLP techniques in the analysis of regulatory speech in these uniform corporate disclosure documents to provide useful insight into left-hand side outcomes, whether industry or filer specific. This small thesis is an initial step in opening the door to a detailed discourse on the informativeness or lack thereof, of regulatory speech within these documents. Originally undertaken in the context of a larger research design seeking to show to opposite, mainly that a nuanced approach to regulatory speech was necessary due anticipated stark differences in the industry and filer differences in regulatory discussion due to applicable regulatory frameworks, this projects seems instead to support the view that regulatory speech in these filings is highly uniform across industries even given the varying regulatory frameworks to which they are subject. Using a simple general regulatory word list paired with a list of words generally indicating responsibility or effect, the simple NLP methodology undertaken here highlights the issues associated with what may be boilerplate or response uniformity in analyzing firm regulatory discussions in the context of annual reporting.

This project analyses 10-K annual reports filed over a 14-year period, 1997 through 2010, employing a paragraph-by-paragraph analysis method to locate and grab paragraphs discussing general regulatory topics and analyze those paragraphs for the number of words indicating effect, implication, and/or responsibility. In short, two naive word lists were created in an effort to identify firm discussion of regulatory implications, burdens, or regulatory “blame” in an initial inspection of filer-wide uniformity in discussions of regulators, laws, and other regulatory regimes which effect, or may be included or discussed by firms as affecting firm productivity, regulatory and or tort risk, among other negative firm implications,

restrictions, or limitations. To this aim, general, non-industry non-regulation specific words were chosen with the intent to get a broad, naive and simple sample for a swift, and easy comparison between industries. General filer-wide summary statistics and temporal changes have been provided, with explanation, as well as a surface level industry analysis and comparison with accompanying explanatory detail for three selected industry subgroups. The comparative subgroups were chosen with specific reference to their regulatory regime differences, obligations, and differences in industry standard and practice. The conclusory bulk of this project suggests several structures and practices with regard to these regulatory report filings which may be at play in creating uniformity in sections devoted to regulatory discussion, boilerplate, or the explanation of risks included in these filings.

Don't Panic.
Douglas Adams, Hitchhiker's Guide to the Galaxy

SECTION I

Brief Literature Review

While there are many applications for and analyses of the words and data included in 10-K EDGAR filings, the study of regulation-oriented speech and disclosures is only a fringe use of the expanded interest in Natural Language Processing (NLP) techniques in corporate, legal, and business research communities. While the Corporate, Finance, Accounting research community have looked at 10-K filings and their varied data and text components as meaningful determinants of left-hand side outcomes for years, regulatory speech and legal regime-oriented speech has been predominantly absent from these analyses. Legal regime reactions and speech outside of corporate or reporting firm outcomes, while perhaps interesting to the legal research community and general Law and Economics studies of corporate disclosure, requirement, and speech; the focus of these research communities thus far has been on corporate political speech following *Citizen's United* (2010) and the complicated theoretical and jurisprudential orientation towards the identity of corporate entities and the rights or obligations with regard to their shareholders or ultimate owners. Specific studies oriented at tying firm discourse on regulators and the legal frameworks to which they are subject within the required disclosure framework is absent in the major corporate and legal research communities. This literature summary begins summarizing the foundational works by Tim Loughran and Bill McDonald in applying NLP tools and techniques to financial reports and corporate arenas and summarizes the current focus of other applicable research regimes within the Finance, Accounting, General Business, and Legal research communities. Currently, there is not a robust foundation of applicable projects from which this project can claim to originate, instead the simple summary provided seeks to explicate possible issues with extending the analysis of legally oriented disclosure language in applications similar to those included in this literature summary.

From the corporate finance, accounting, and general business perspective Tim Loughran and Bill McDonald are the fathers of the application of Natural Language Processing (NLP) techniques in finance arenas, reporting, and analysis. Their projects include, but are hardly limited to; studies detailing the

readability of corporate disclosures using traditional semantic indexes, the mislabeling of positivity and negative sentiment in word list applications in financial reports, and several summaries and discourses on the appropriateness, limitations, and applicability of word list usage and NLP techniques in finance and accounting (Loughran and McDonald, 2009, 2011, 2014, 2015, 2016, 2017, and forthcoming). While neither of the authors would likely approve of the wordlist herein employed for encumberment or burdening as “Blame Words” due to their extensive explanations regarding the inappropriateness of using negative colloquial tone indicators in corporate disclosures, their specific works in the applicability of word lists and textual analysis generally and in the fields of finance and accounting provided the foundation for the structure of this research design (see, Loughran and McDonald, 2015 and 2016). Their applicable works focus on applying NLP techniques to 10-K filings through 1994 - 2007 and 2009, in detailing various implications in speech, rhetoric, and the applicability of plain English tonal indicators and their appropriateness with regard to financial behavior, market movement or status, and firm outcomes. In short, the authors laid the foundation for applying NLP methodologies and orientation to left-hand side outcomes in the context of corporate disclosures and financial reporting speech. In their 2016 summary, *Textual Analysis in Accounting and Finance: A Survey*, the authors detail useful guidelines and meaningful limitations on the usage of these NLP techniques in an effort to highlight the appropriate orientation of coming authors when using such flexible, and sometimes opaque techniques when making left-hand side suggestions about firm, market, or sentimental outcomes resulting from NLP, word list, word counting, and other NLP based data gathering methods (Loughran and McDonald, 2016). This project, while highlighting the lack of supplemental foundational orientations on law or legal regime specific applications of the Loughran and McDonald methodologies, seeks to provide another highlight suggesting that results oriented at tying left-hand side outcomes to legal or regulatory speech markers provided by NLP methods may require nuanced or supplemented support as Loughran and McDonald have themselves suggested regarding their more general applications.

Specific academic works citing, using, and guided by Loughran and McDonald’s fundamental principles and application methodologies in the Accounting, Finance, and General Business and

Organizational fields have generally focused on tying left-hand side outcomes in financial performance, market orientation, small investor behavior, and other corporate or topically specific data points (like management turn over, governance measures, organizational design, and specific litigation) to language included in corporate disclosures, including, but not limited to, Forms 10-K, 10-Q, and 8-K (a detailed list of SEC Forms and Form usage can be found on the SEC website, link included in the Bibliography section)(summaries included, Loughran and McDonald, 2016; Li, 2010; Fisher, et al. 2016; and Kearney and Liu, 2014). While many works related to or arguing similarly to Loughran and McDonald's works analyzing textual tone (Jegadeesh 2013), text usefulness to readers (Frankel, et al., 2017, using NLP techniques on conference calls and predictions), and text readability (Li, 2008; also using the FOG index readability measure); others have detailed the textual changes in regulatory filings over the years in length, complexity, and detail (Dyer, et al. 2017; Cazier and Pfeiffer, 2016; among others), and still others have used NLP techniques to study; litigation risks and disclosure components (Nelson and Pritchard, 2007); risk factor disclosures (Campbell, et al., 2014); and financial statement fraud indicators (Goel and Gangolly, 2012). Tangential in argument to the initial call of this project, in seeking to explicate legal analyses surrounding annual reporting, several works have sought, with specific respect to the environmental reporting, standards, and disclosures, to detail the need for explicated and specific frameworks for such disclosures and reporting compliance, a similar argument posed in the concluding remarks to this project's introductory sections (Peters and Romi, 2013; Eccles, et al. 2012). While one final vein of corporate research seeks to elucidate which topics, should be of interest to investors (Starks, 2009), an additional component pertinent to the selected introductory discussion. Such prior studies and methodologies, while instructing the use of word lists, NLP techniques, and orientation employed in this project, have yet to breach employing these techniques in analyzing regulation, legal regime specific speech.

Legal research pertinent to the subject of corporate discourse on laws, regulation, and corporate speech overall, is scant in the arena of required disclosures. Instead, legal research on corporate speech can be categorized into two separate topical categories; corporate political speech as it relates to the first amendment (pre and post *Citizens United* (2010)), and corporate speech as it related to preferable treatment

under the laws resulting from campaign contributions and corporate political activism. Corporate personhood and jurisprudential topics regarding corporate or firm rights to free speech can be arranged as those preceding the 2010 U.S. Supreme Court case *Citizens United (2010)* and those following or arguing its validity in relation to the rights of associated parties, with respect to shareholder, manager, director, and overall corporate entity speech freedoms (Redish and Wasserman, 1997-1998; Siebecker, 2006-2007; Winkler, 2006-2007; Bebchuk and Jackson, 2010; Tucker 201-2011; Melone, 2010-2011;). While additional legal discourses, discuss the validity of restricting corporate speech or curtailing statement forms by the SEC (Page and Yang, 2005-2006), the overarching legitimacy of enforcing Regulation Fair Disclosure (Humes, 2006-2007), and other non-disclosure related reporting obligations and duties arising under U.S. Federal Securities Laws (Backer, 2003). However, firm speech discussing regulators, laws, or other regulatory regimes or risks in corporate disclosure documentation has remained untouched by prior legal research, a fact which motivated the initial design of this inquiry into firm legal and regulatory speech in form 10-K.

10-K Reporting Requirement and Components

Not all companies are required to submit annual reports to the Commission's EDGAR filing database, but those which have undergone their Initial Public Offering (IPO) and have a volume of shareholders numerous enough to be considered "public" or "publicly traded" under the applicable Securities Acts, (defined in greater detail below) definitions are required, under the Acts, to submit annual and sub-annual reports to the Commission. These annual 10-K filings are overseen by the Division of Corporate Finance (DCF) which oversees the filings and registrations required under the Securities Act of 1933, defining the various securities instruments, their sale, and all relevant disclosures and sale reporting and retail, involved in securities promotion in the United States; the Securities Exchange Act of 1934 defining the transaction of such publicly, privately, and or off market traded securities, exchanges, over the counter retailers, and all initial offerors; and the Sarbanes-Oxely Act of 2002 and Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 which include additional definitional, application, and

reporting guidelines to which these publicly traded firms must adhere.* Generally speaking, DCF office seeks to ensure that investors are provided with the information, disclosures, and reports required to make informed decisions on publicly listed and traded companies, stocks, and other securities instruments offering and circulated in the U.S., additional information on the calling, duties, and powers assigned to the DCF office can be found on their webpage, but their divisional agency description reflects the below:

“In support of the Commission’s mission to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation, the Division of Corporation Finance seeks to ensure that investors are provided with material information in order to make informed investment decisions, both when a company initially offers its securities to the public and on an ongoing basis as it continues to give information to the marketplace. The Division also provides interpretive assistance to companies with respect to SEC rules and forms and makes recommendations to the Commission regarding new rules and revisions to existing rules.” *Taken from the DCF homepage, Last Modified: Jan. 31, 2017.*

For the purposes of this project the key facets of the DCF and over all Commission reporting regime is the concerted focus on report uniformity. The Commission, through the DCF, its website, and the various Commission Accounting Offices provide a wide supplement of practices aids, reporting submission guides and guidelines, and other financial reporting and filer oriented structural guides and definitional riders and references. In short, the Commission provides a wide panoply of guides for filers in an effort to enforce relative uniformity in information submitted and report structures employed by firms issuing the required annual filings. What follows here, preceding a description of the data herein included, is a brief summary of the detailed informational guides, sample forms, and informational sections suggested by the Commission for all 10-K filing firms.

Informational Obligations

The annual 10-K report functions as the annual report, supplemental to the required annual report issued directly to shareholders, which is typically much more detailed, relaxed, and also typically written in the company’s tone and presented in concert with the required annual meeting where shareholders elect directors and handle other applicable shareholder and board functions. The 10-K form, unlike the annual

* Links and citations to relevant SEC materials, practice guides, and other material used in this requirement summary can be found in the Bibliography section at rear.

shareholder report, is formulaic and provides a scripted, but comprehensive overview of the company's business, financial, and other material conditions, as prescribed by the reporting requirements binding all U.S. issuers of publicly traded securities under the 1934 Act. In the Commission's provided sample form instructions, the "Rules as to Use of Form 10-K" specifically cite the applicable 1934 Act reporting requirements the Form is designed to meet and for what purposes and covering what required annual disclosures the Form is intended to cover under the Section 13 and or 15(d) obligations of the 1934 Act (15 U.S.C. 78m or 78o(d)). Which references obligations under the Act as follows;

"Every issuer of a security registered pursuant [to this Act title] shall file with the Commission, in accordance with such rules and regulations as the Commission may prescribe as necessary or appropriate for the proper protection of investors and to insure fair dealing in the security...such annual report (and such copies thereof), certified if required by the rules and regulations of the Commission by independent public accountants, and such quarterly reports (and such copies thereof), as the Commission may prescribe." 15 U.S.C. 78m(a) and 15 U.S.C. 78m(a)(2) respectively. Periodic Citation, Cornell Law, <https://www.law.cornell.edu/uscode/text/15/78m>.

While the 10-Q and other non-annual filings, like the 8-K filing for material events, are more brief, and often take a slightly more free form often providing for the addition of appendage of supplemental forms and materials, the bulk of the 10-K is meant to provide a regulation based summary form for the annual wellbeing of reporting obligated firms in a form and structure which is designed, without flourish, to inform investors of informational parameters and through a form designed to meet the obligated disclosure goals and guidelines provided in the 1934 Act and adhering to the Commission's calling to keep investors well informed.

Information Included

The Commission's calling is not only to provide investors with meaningful or material information regarding the securities and companies into which they may invest, but to do so in a way that, when a consumer sees a portfolio of various different stock offerings, whether those already listed, or those as several IPO offerors coming to market, the investor can compare the various investment opportunities using a uniform scale or material measures form, no matter if the issuers are from very different industries. TO this end, the Commission has designed a sample 10-K Form, for use by filers to maintain relative uniformity

both in structure, but also in an effort to create comparability in the material corporate information, accounting reports, and financial statements included. While the Form appears fill in the blank, the actual 10-K filing is done free form and submitted by part through the online Commission repository EDGAR (Electronic Data Gathering, Analysis, and Retrieval system) provided through registered access with the Commission via their agency website. The Commission provides the following directives on completing the Form for filers using the Sample Form 10-K as a guide:

“This form is not to be used as a blank form to be filled in, but only as a guide in the preparation of the report on paper meeting the requirements of Rule 12b-12. Except as provided in General Instruction G, the answers to the items shall be prepared in the manner specified in Rule 12b-13.” [*FORM 10-K ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934, General Instructions, <https://www.sec.gov/about/forms/form10-k.pdf>, Rule 12b references timeliness guidelines dependent on Filer Size as prescribed under the 1934 Act as amended*]

The Form itself then details four subparts which included within-section subparts for applicable material information to be included, detailed, and explained. [See pages 8-11 on the Sample Form]. The sections lay out the filing formula, or required sectional inclusions deemed informationally material of necessary by the Commission, including information on company history, organizational structures, equity plans, subsidiaries, executive compensation, and relevant financial information and statements. A greater explication of the information sought in 10-K filings is detailed in the following section. To this end the Commission provides supplemental guidelines and practice handbooks to report authors specifying the specific labels and information components to be included in each section. These guidelines include both specific guides for the 10-K report as well as a full Financial Reporting Handbook, which provides equivalent detailed guidance on each information submission required or suggested by the Commission for all public or otherwise registered companies. These guides provide a highly formulaic and very detailed view of how and in what format and structure information for filers under the 1934 Act should compile, report, and display their information for investor consumption. The formulaic nature of the partitions and subsections is highlighted by the Commissions investor guidance on “How to Read a 10-K,” provided on the Commission website, detailing the information included, expected, and required within each of the Commission’s defined sections.

The 10-K filing system has been designed, to this end, to ensure that investors have accessible uniformity in the information they are being presented with regard to each possible offered securities investment available to them in the U.S. market, meaning firms from different industries are funneled into a uniform filing system oriented at

providing a uniform and consistent presentation of material firm information. This intended uniformity, as will be detailed in this short project below, is highlighted.

Uniformity for Comparison

While clear from some quotations highlighted above, the Commission requires, structures, and provides practice and submission guidelines for 10-K filings so that their contents can be uniformly compared across firms, industries, and operating characteristics. However, while the reporting guidelines speak to more than just creating uniformity in report structure alone, their components imply that there is a specific subset of firm related information appropriate or applicable to investor decision making and by extension firm performance as determined by the Commission. The 10-K is divided into sections by topic and applicable headers for each discussion of required on all 10-K forms submitted. The 10-K is broken down into four sections, each “Part” having to do with a particular granularity of overall firm performance. A brief title list of the Parts and subpart or “Items” is provided below, regulatory or legally relevant Item text summaries have been included, courtesy of the Commission’s “How to Read a 10-K” Investor guide. Summary text sections have been included and emphasis has been added to Item sections and subtopics where companies tend to discuss legal issues, regulation, regulators, and topics including legal words picked up by the NLP techniques used in this project.

PART I

Item 1 - “Business” requires a description of the company’s business, including its main products and services, what subsidiaries it owns, and what markets it operates in. This section may also include information about recent events, competition the company faces, *regulations that apply to it*, labor issues, special operating costs, or seasonal factors. This is a good place to start to understand how the company operates.

Item 1A - “Risk Factors” includes information about the most significant risks that apply to the company or to its securities... Some risks may be true for the *entire economy*, some may apply only to the *company’s industry sector or geographic region*, and some may be unique to the company.

Item 1B - “Unresolved Staff Comments”.

Item 2 - “Properties”.

Item 3 - “Legal Proceedings” requires the company to include information about significant *pending lawsuits or other legal proceedings, other than ordinary litigation*.

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”

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 Accountant Fees and Services”

PART IV

Item 15 - “Exhibits, Financial Statement Schedules”

At first glance the first point of notice is that there are no sections, parts, or items dedicated solely to the discussion of applicable legal regimes, regulations, or other regulatory frameworks. Instead, topics regarding applicable legal regimes and regulatory implications arise as organic components of sections dedicated to general business discussions, risk factors, and pending or ongoing litigation. In short, if the Commission’s suggested sectional components and requested descriptions are taken as indicative of components meaningful to investor comparison, applicable legal regimes, regulators, and frameworks may be viewed as unimportant to overall corporate performance or the appropriateness or future viability of any investment in a reporting public company. More plainly, where discussion of legal concepts, regulators, or applicable legal regimes arises, it is not meant to be used as a comparative measure made important to investor purchase decisions, but instead an outgrowth of firm discussion of general business summary, overarching regulatory risk, or an instance of specific litigation. Hence, discussion of legal topics is not a topical discussion purposefully elicited for the purpose of making a firm-to-firm comparison and instead arises as a descriptive component included by choice without any specific uniform structural request from the Commission or any specific applicable guidelines for the inclusion of such legal regime components. Without applicable structures oriented at providing required uniformity in disclosure structure and components for specific sectional discussions of legal and regulatory issues, employing the language used by firms to discuss those topics presents structural issues this simple project seeks to highlight.

Initial Concluding Remarks

To conclude these introductory and literary background sections before moving on into the simple sample and data comparison included in this project, the summary of Commission specific informational component requests and structural 10-K components was included to highlight the lack of legal or regulatory discussion elements formally required by the Commission and hence the structural lack of comparative guidelines through which meaningful uniform comparison of firm speech on legal or regulatory topics can be conducted. Where differences in the language

used to discuss required disclosure components, like financial performance, executive compensation, or other major “Items” may lead to meaningful outcome or left-hand side outcomes, without specific calls to submit legal or regulatory discussion or meaningful bounds through which such discourse should be submitted creating meaningful marks of comparison, comparing firm outcomes based on such legal and regulatory speech should be undertaken carefully.

When initially undertaken, this project sought to draw comparative delineations between firms based on their discussion of regulators and applicable legal frameworks. However, in the initial naive sample that follows, demarcation or meaningful differences between firm speech on regulators and applicable legal frameworks is sparse and a comparison of three industry groups is used to highlight this interesting outcome. In hopes of motivated further research on how disclosure burdened firms discuss regulators, this project concludes with suggested reasons why firm regulatory speech in the general business summary, risk, and legal proceedings sections may be either too sparse or too similarly bland to provide meaningful comparison for left-hand side outcomes.

The chances of finding out what's really going on in the universe are so remote, the only thing to do is hang the sense of it and keep yourself occupied.

Douglas Adams

SECTION II

Data Summary

This project covers 10-K forms filed between the years of 1997 and 2010, including 14 separate years of filings made to the Commission's EDGAR repository totaling more than 230,000 total 10-K's. Each file was cleaned prior to the implementation of the NLP methodology detailed below using the BeautifulSoup function for files not already in raw text form, those necessitating a conversion from raw HTML formats. Each year saw roughly between 12,000 and 17,500 filings, with the length of filings, measured by number of total paragraphs roughly increasing through the sample and the total number of filings submitted each year slowly decreasing from 1997 through 2010. A summary of all filer data and a three-industry group comparison were undertaken to compare the word counts found and overall filer and file characteristics found within the sample. Each filing set was arranged by file year.

NLP Program

The Natural Language Processing (NLP) methodology used in this project employs a simple word counting mechanism and word identification through two small operating word lists. Not unlike Loughran and McDonald's works on textual sentiment using positive and negative word lists and accumulating word counts for the various tagged words found in text, the program created for this project located regulatory or law oriented words from a small word list of general legal terms and once located, then stored the paragraph of text in which the word was located to a running string vector and then, at the end of the text document searched that string of compiled regulatory paragraphs for "blame" oriented words indicating responsibility, burden, or "due to" effects again, counting the number of "blame" oriented words found within those regulation or law specific paragraphs. The counting mechanisms and coding was done through Python and using the Natural Language Tool Kit (NLTK) plug ins and commands. The word counting mechanism and paragraph counting designs in this project are very simple and the author hopes that these techniques and the paragraph-by-paragraph orientation at 10-K topical speech can be used in other projects by future researchers or those who, like the author, are still developing their programming skills.

The program has been included at rear in PDF form in the appendices for reference, but in brief summary of the programmatic components, a process has been included below as well as the two applicable word lists used in word counting. The program begins by initializing and downloading the necessary packages, script sets, and operating functions to compile the word lists, execute word counting, and create the data output file once the word counts and paragraph counts, and other characterizing data points have been extricated from the files. In the first step, the program creates two vectors of words for the word counting mechanism to use when parsing, line by line, through the text file by importing the “Regulatory Words” and the “Blame Words” from their respective csv files using the methodology and structures used by Loughran and McDonald in their word counting projects. Each list of words, provide below, is turned into a vector, through which the word counting mechanism will loop in checking each line and each word element of each line as the program reads through the document in step two. Each word is a raw csv file listing the words, by structure they are listed in all caps, but since this program operates without specifically calling to capitalized words, stemming, or removing stop words, the words were left un-stemmed, and each string of paragrammatical pulls taken once a regulatory word was located was taking in full, with only punctuation being removed.

Regulatory Words:	ACT	Blame Words:	BAN
	AUTHORITIES		CAUSE
	AUTHORITY		CONTROL
	GOVERN		DUE TO
	LAW		EFFECT
	LEGAL		IMPACT
	LEGISLAT		IMPOSE
	ORDINANCE		LIMIT
	POLICIES		OBLIGAT
	POLICY		PROHIBIT
	REGIME		REGULATE
	REGULATION		REQUIRE
	REGULATOR		RESTRICT
	RULE		
	STANDARDS		
	TREATIES		
	TREATY		

Once the vectors have been imported into vector form, the program is then sent year-by-year into the grouped filings, opening each document one-by-one and reading each text file line-by-line. Initial lines

containing filer-firm identifying information like unique firm identifier (CIK, Central Index Key number), firm name in text form, filing time and date, and industry identifier (SIC, Standard Industrial Classification number) are extricated and retained for identifying inclusion in the later data line which is appended at the end of the file reading process. The program counts each paragraph using the new line characters denoting empty space in the text file, and when a regulatory word is found within a paragraph, the lines of text making up that paragraph are saved to a repository vector in string form, once the reading functions comes to the end of the text document for that 10-K file, the program begins with step three. In step three, the string vector of regulatory paragraphs, paragraphs which were flagged as containing regulatory words, once punctuation is removed, is then counted for “Blame Words” in the same looping word counting mechanism that located the regulatory words.

Finally, in step four, the key demographic and identifying information recorded within the program for that particular filings are converted into a single line of comma separated text, and appended to the running vector of output data lines for that filing year, and once the entire year’s filings have been read, the comma separated line are converted into a csv file of output data. A sample output data file is supplied at rear in the appendices. This simple procedure returns equally simple, but enlightening data as to the general inclusion of regulation and legal regime-oriented speech which has been separated into the data presentations below for both a summary of data across all filers and a sub-selection of industry participants based on three separated industry identified subgroups.

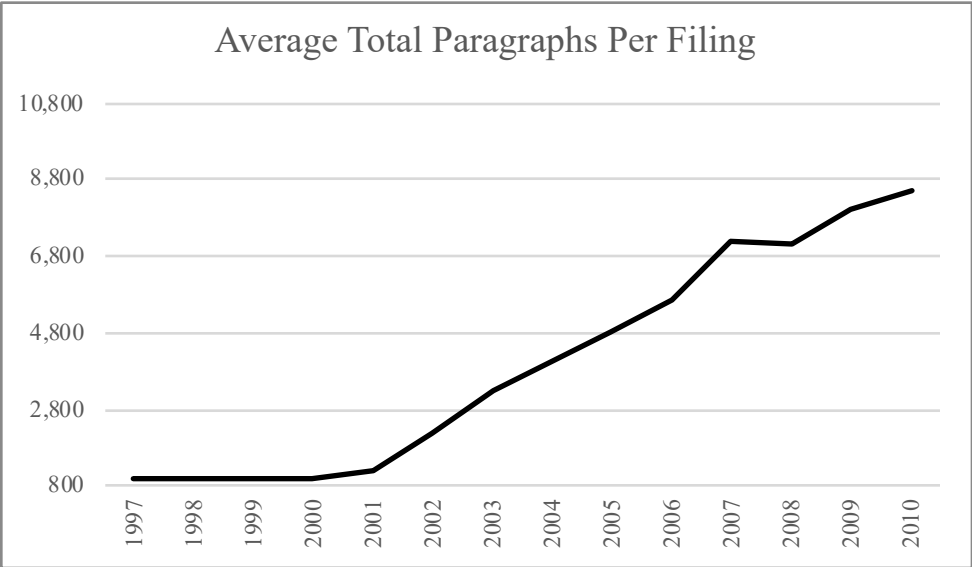
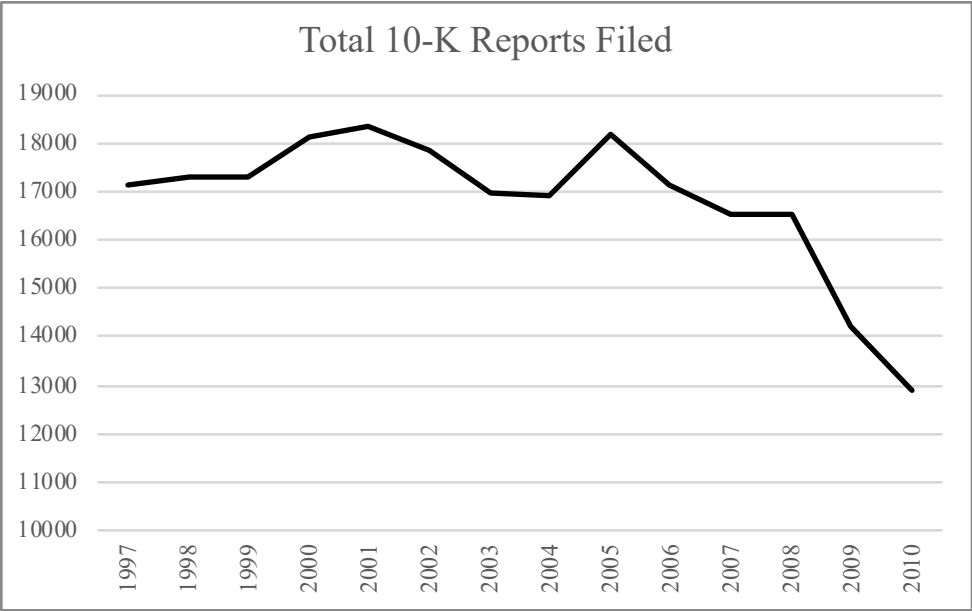
I refuse to answer that question on the grounds that I don't know the answer.
Douglas Adams

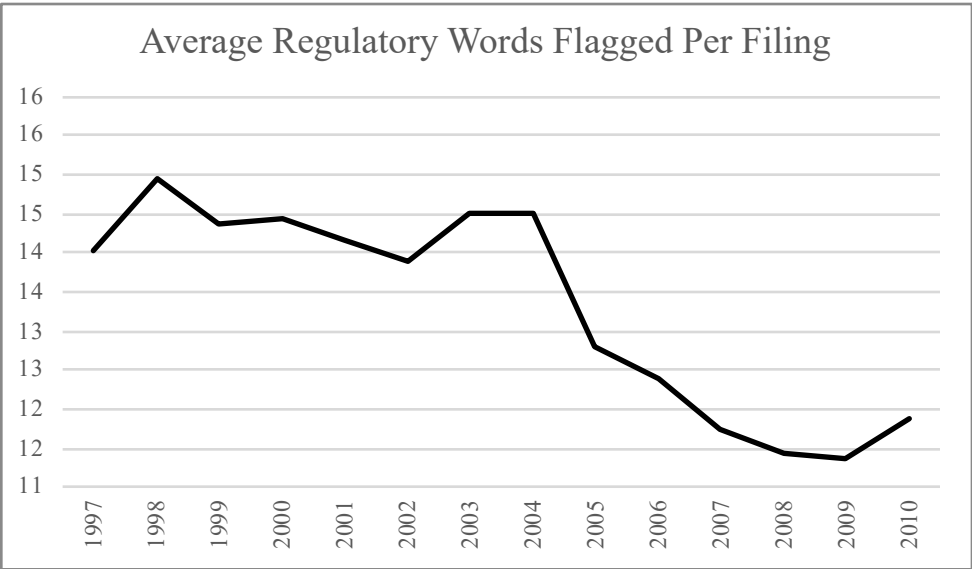
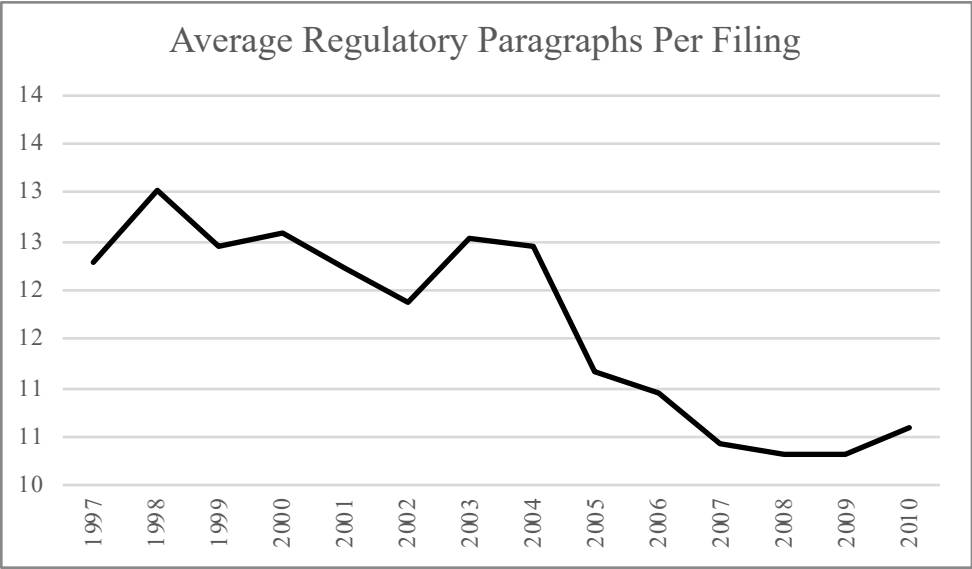
SECTION III

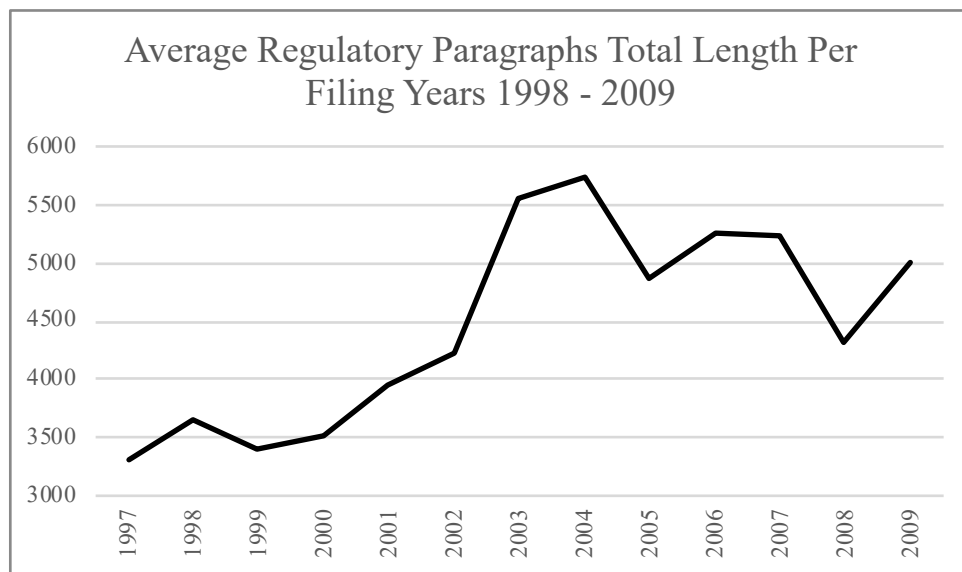
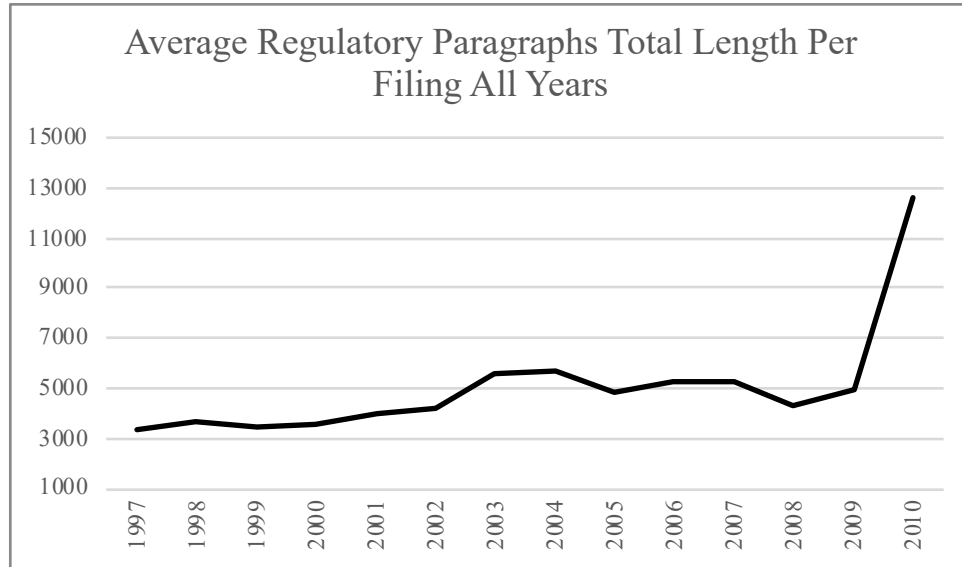
All Filers Data and Initial Summary Remarks

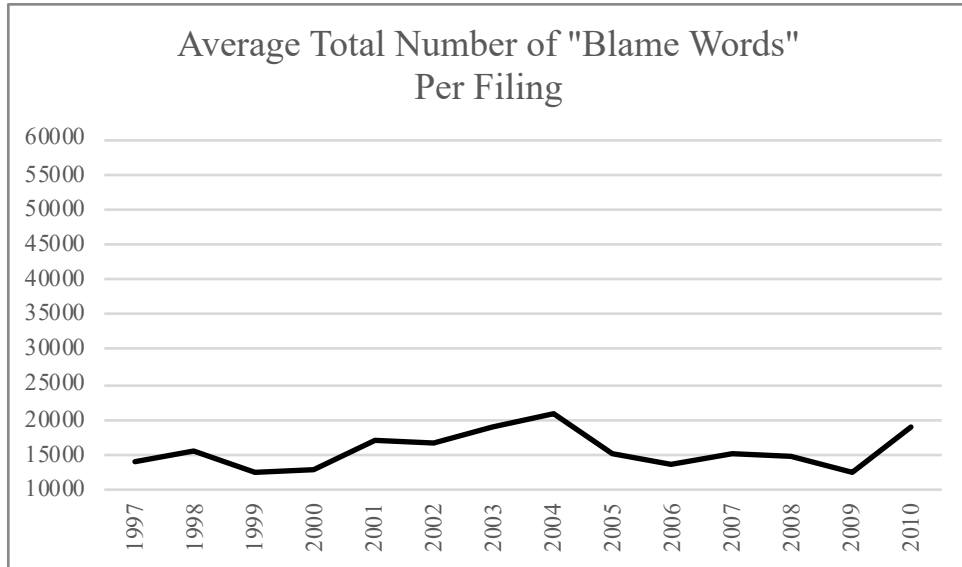
Year	Total Filings	Average Paragraphs	Average Reg. Paragraphs	Average Number Reg Words Flag'd	Average Reg Paragraphs Total Words	Average "Blame" words
1997	17147	978.590016	12.2764332	14.04047355	3311.627865	13849.007
1998	17,283	1002.14951	13.0299716	14.9572412	3645.08604	15528.147
1999	17,273	989.625601	12.4566665	14.377294	3414.08481	12449.9
2000	18,135	1006.57541	12.5793769	14.4310449	3524.18875	12905.304
2001	18,317	1192.12191	12.2279303	14.1599607	3953.7813	17043.384
2002	17,825	2164.47961	11.8808976	13.8873492	4217.01532	16762.141
2003	16,981	3276.2395	12.546022	14.5142807	5549.9377	18912.36
2004	16,914	4046.30708	12.4619842	14.4949155	5728.47611	20745.635
2005	18,195	4844.98873	11.1621325	12.799945	4876.16697	15077.44
2006	17,154	5680.97196	10.9470677	12.3952431	5243.86365	13773.917
2007	16,504	7220.05284	10.4194135	11.7424261	5225.68317	15129.344
2008	16,543	7116.11243	10.3126398	11.4379496	4319.48709	14746.89
2009	14,226	8018.58864	10.3195557	11.3812737	4993.82602	12271.789
2010	12,908	8518.42268	10.5868454	11.8703905	12627.2261	19017.428

While the total number of reports filed follows a roughly decreasing pattern from 1997 – 2010, the average length as measured by the total number of paragraphs included in each filing steadily grows throughout the sample. Figures have been provided on the following page. While data was initially taken from 1997 – 2012, errors in the programmatic application of the NLP methodology led to a distinct change in data grabbing following 2010; this could be the result of applicable reporting guideline and regulatory changes relating to public filers updated by the Commission in 2010 changing the applicability of pertinent word usage and the appropriateness of the word grabbing mechanism used, however, analysis of these pattern responses is outside the purview of this small project and therefore was simplified and removed until a sample of post 2010 filings can be undertaken using a different grabbing technique. However, the general summary progression of regulatory and blame word counts supplied in the diagrams that follow provides an interesting and discourse motivating snapshot of the general filer regulatory discourse included in these filings for summary purposes.









Time is an illusion. Lunchtime doubly so.
Douglas Adams

SECTION IV

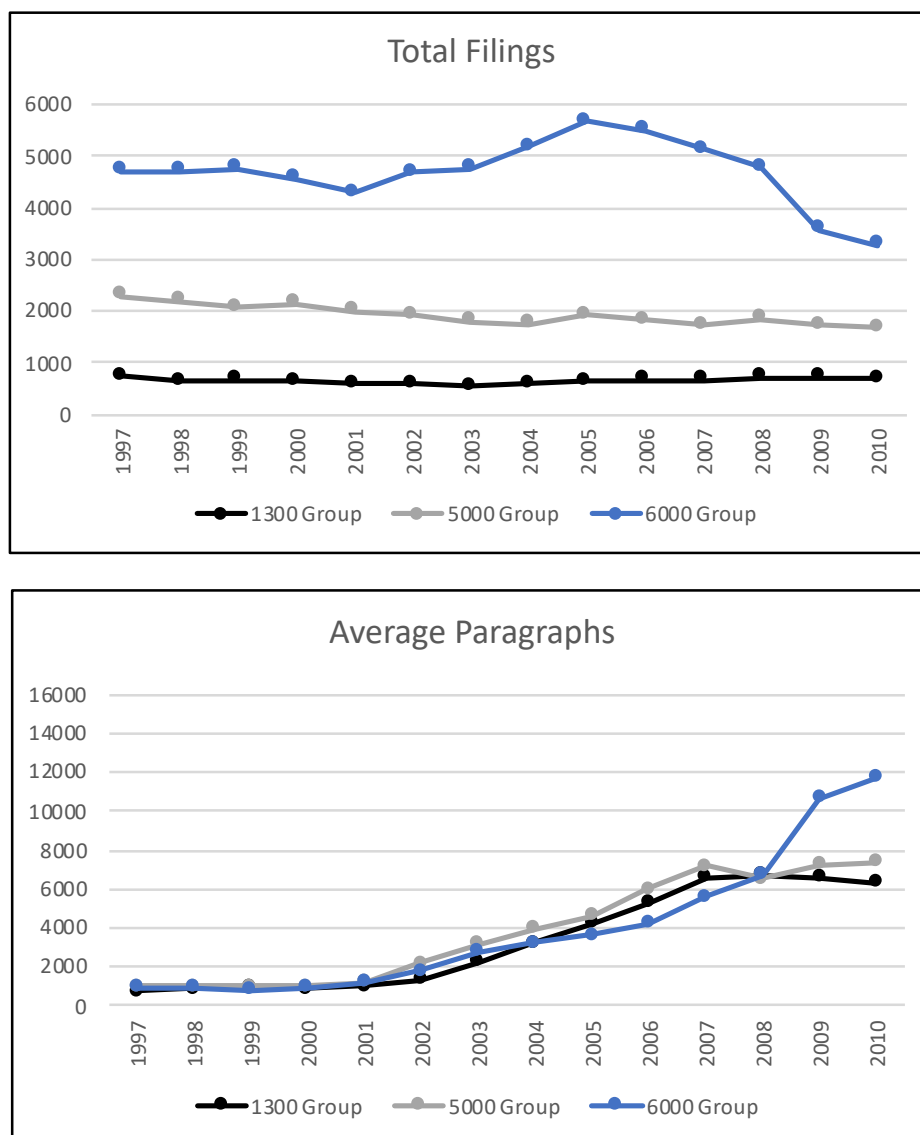
Industry Comparison

While the summary figures previously shown show an interesting overtime development over the average and aggregate filing characteristics for U.S. reporting public companies, the focus of this paper is instead on a three-industry comparison meant to illustrate the difficulties in using regulatory or legal regime oriented speech in the context of 10-K filings. The Commission classifies reporting companies into industry groups or segments using the SIC or Standard Industrial Classification system.

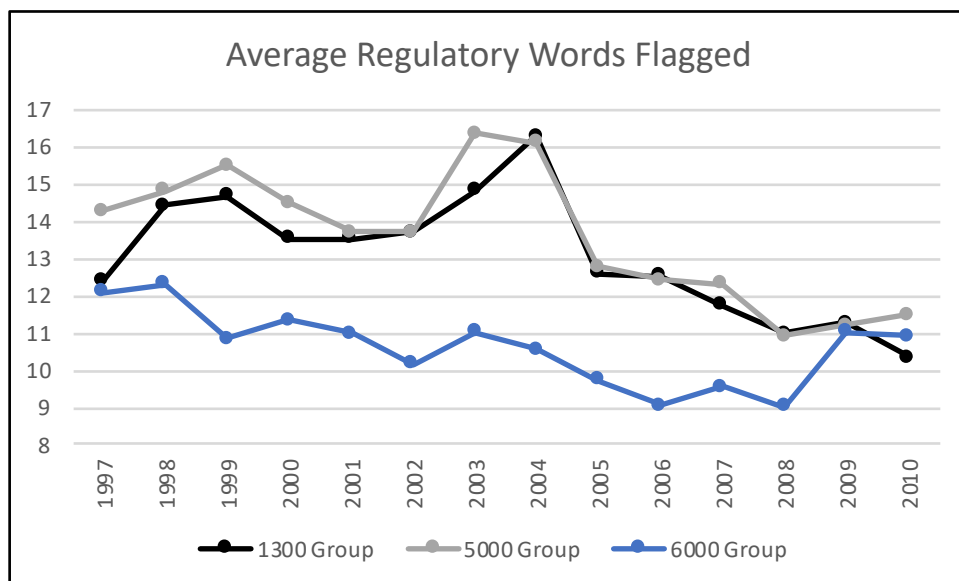
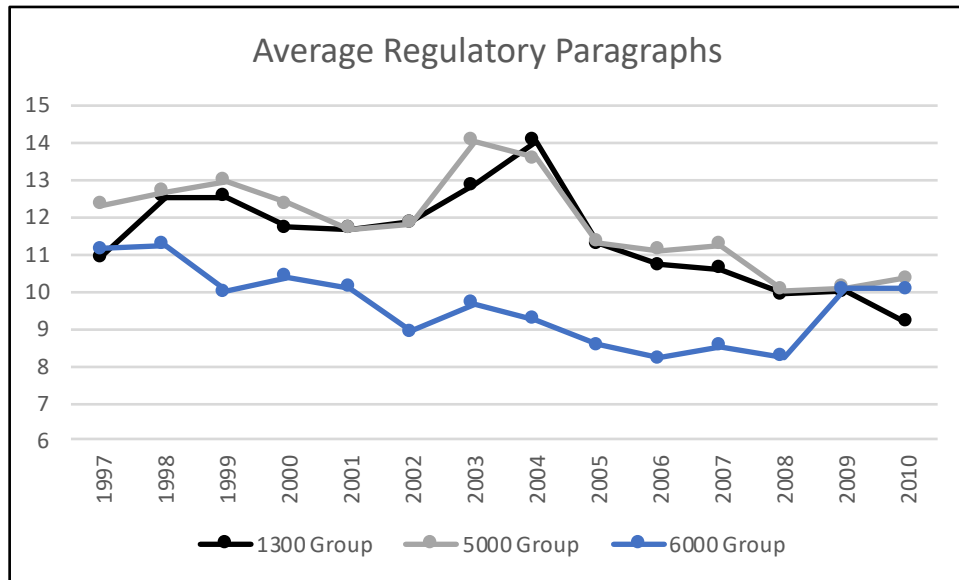
The Standard Industrial Classification Codes that appear in a company's disseminated EDGAR filings indicate the company's type of business. These codes are also used in the Division of Corporation Finance as a basis for assigning review responsibility for the company's filings. For example, a company whose business was Metal Mining (SIC 1000) would have its filings reviewed by staffers in A/D Office 9. SEC definitional description provided on the Commission's webpage *Division of Corporation Finance: Standard Industrial Classification (SIC) Code List*.

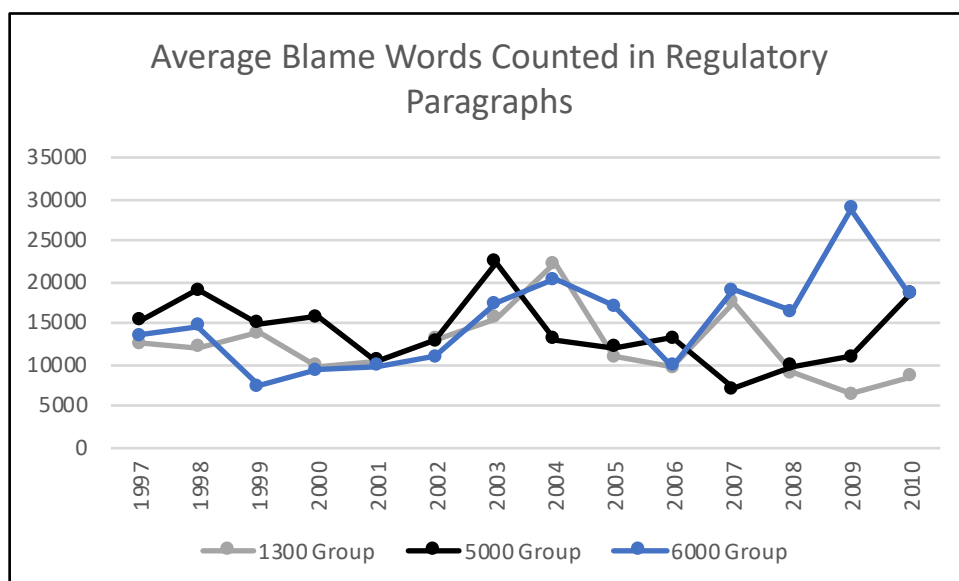
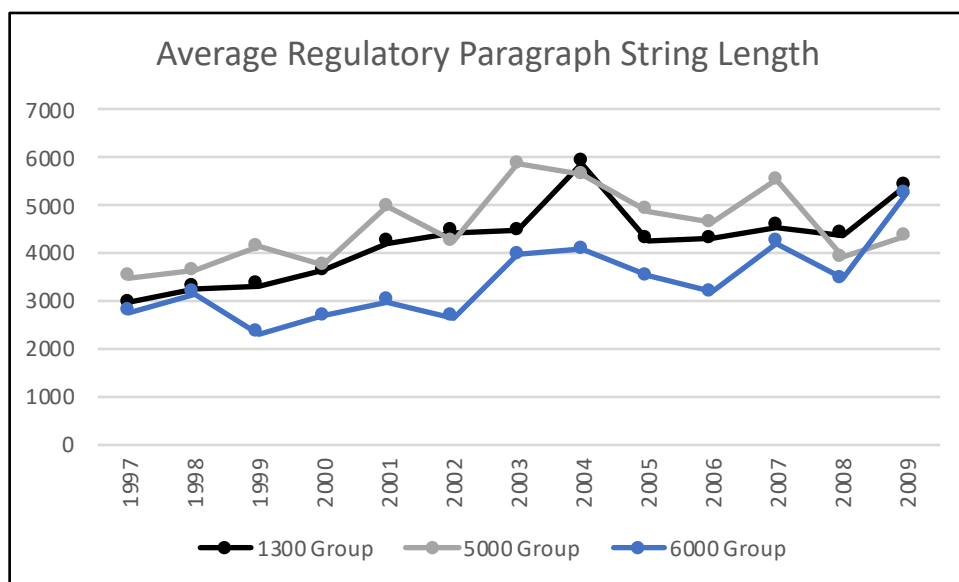
Using the Commission's SIC classifications, three groups of report filers were selected for comparison, Petroleum and Natural Gas companies, Wholesale and Retail companies, and Financial and Investment Institutions, SIC codes: 1311 – 1389, 5000 – 5990, and 6021 – 6282 respectively. These selections were made to highlight the differences in the patterns of legal and regulatory speech and blame words found given the stark differences in the regulatory regimes, frameworks, and compliance standards applicable to firms in each of the three SIC groupings. Petroleum and Natural Gas companies must practice constant compliance with and adherence to strict and detailed state, federal, and international guidelines for products, implementation, and practices in an impact or specific control framework of regulators and regulations. Wholesale and Retail providers, on the other hand, face little structural control over operations and instead face the majority of their regulation in consumer suit, tort and tort reform, and other state and federal guidelines over market pricing and consumer or employee concerns. And the final category, Financial and Investment Institutions, faces a detailed system of investor and consumer protection and market protection regimes detailing strict compliance, disclosure, and reporting guidelines, predominantly focused on entities like the FDIC, the SEC, Finra, and other report and disclosure oriented regulatory

regimes. The regulatory structural components and adherence frameworks present in each of these regimes was initially hypothesized as providing ample motivation for firms to produce varied usage and discussion of legal terms, regulatory speech, and comment on the implications or “blame” words oriented at pertinent regulators. What seems instead to result, is a more nuanced set of regulatory response difference, where firm group trends, while differencing slightly, do not seem to reveal marked differences in usage and blame.**



** A full table of data used in the three SIC group comparison is provided at rear in Appendix 4. 1300 Group: Petroleum and Natural Gas, 5000 Group: Wholesale and Retail, 6000 Group: Financial and Investment Institutions





While each sector shows a marked difference in the number of filers, the relative behavior of each of the other groups shows a more nuanced, and to some degree, unexpected differences in the number of regulatory words flagged, and similarities in the length of regulatory strings reviewed and in the number of blame words used is, again, surprising, given the differences in regulatory frameworks applicable. While increased reporting guidelines passed in 2008 may be cause for the explosion in the length of Financial and

Investment Institution filings following 2008, the constant deficit in general regulatory speech, indicated in regulatory words flagged and in regulatory paragraphs overall, is relatively surprising. In short, this simple comparison demonstrates the need for a nuanced approach to using legal or regulation oriented speech as a meaningful left-hand side indicator in the context of 10-K filings. While words used pertaining to regulatory concepts and legal issues may lead to within industry interest, for example, as a motivating component for greater research into the difference in form length between industries following the financial regulations instituted as effective in 2008, using regulatory speech as an overall indicator in tying firm outcomes may not provide meaningful reflection on actual applicable legal regimes. Hence, this project seeks to provide this brief industry comparison to highlight the difficulty in comparing these legal and regulatory mentions where this language and these general regulator words are not, by the structure of the information requested, guided, or compelled for inclusion in these reports.

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

Douglas Adams, Hitchhiker's Guide to the Galaxy

SECTION V

Concluding Remarks

While other projects have tied left-hand side outcomes to language contained in 10-K reports, using the legal and regulatory oriented speech in these reports in the same way may provide difficult or meaningless without implementing a nuanced, sector-by-sector, industry-by-industry, or some more specific approach. Without a specific call or uniform guidelines for the inclusion of such regulatory information or discussion for means of comparison, like other sections included in these filings, using regulatory speech for such purposes in supposing ties to left-hand side outcomes may be unwarranted, difficult, or may fail to provide statistical significance. While this outcome has been disheartening to the author, some minor further discussion below as to why this particular issue may be the case below may help to further other research paths oriented at dissecting this type of report information.

In discussion with several lawyers who file Form 10-K on behalf of their clients, several possible reasons for these nuance issues arose, if, unlike tying such legal discussion to left-hand side outcomes, the study of these particular topics is undertaken in the context of these filings, perhaps those topical focuses, could yield a more reasonable, but less NLP focused output data of interest. For many of these filings, these legal references surfaced in boilerplate language detailing possible risks, especially those in the case of foreign or other applicable legal regimes. For many filing attorneys and firms, the use of stock boilerplate or copy boilerplate language is common, if not the norm, and should a small firm lawyer be filing the same, finding appropriate language by using boiler plate already included in a prior filing by a third party is not uncommon, it is quite the case that copying is flattery, and best practice to some regard. The same seems to hold true for the operating inclusion of legal disclaimers. Recycling disclaimer and boilerplate language both within firm and filer to filer is common, lawyers, not unlike the author, feel there is hardly any need to reinvent the wheel, and as the old adage goes, it is the non-conforming nail that receives the hammer after all. Finally, industry standard for disclosure, risk boilerplate, and the discussion of risks, regimes, and legal doctrines may also be at fault for these nuances. Smaller public companies may follow their well-

received larger industry bedfellows in conforming to reporting regularity when it comes to boilerplate language, disclaimers, and risk summaries, for the simple reason that, where all companies report identically in such matters, an operating industry standard seemingly results. In short, such behavioral keys in the law firms and accounting departments issuing these reports, while fruit for specific projects on these concepts, muddies the water in using the legal and regulatory comments included in these reports as meaningful identifiers for left-hand side outcomes.

I love deadlines. I love the whooshing noise they make as they go by.
Douglas Adams

BIBLIOGRAPHY

Bibliographical information included below has been grouped by topic, research field, and provider.

Useful Securities and Exchange Commission Provided Links, Materials, and Guides:

ECFR for 1934 Act: <https://www.ecfr.gov/cgi-bin/text-idx?amp;node=17:4.0.1.1.1&rgn=div5>

Landing Page for Sec Corporate Finance Division: <https://www.sec.gov/page/corpfin-section-landing>

Rules and Schedules Guide by SEC: <https://www.sec.gov/divisions/corpfin/ecfrlinks.shtml>

Quick Information listing 10-K by SEC: <https://www.sec.gov/fast-answers/answers-form10khtm.html>

SEC Form 10-K Sample PDF with Sample form and Instructions for Filers:
<https://www.sec.gov/about/forms/form10-k.pdf>

“How to Read a 10-K” Investor Guide by SEC:
<https://www.sec.gov/fast-answers/answersreada10khtm.html>

SEC Practice Filers Manual: <https://www.sec.gov/corpfin/cf-manual>

SEC Forms List: <https://www.sec.gov/forms>

Resources and Citations by Tom Loughran and Bill McDonald:

Link to Notre Dame repository maintained by Bill McDonald: <https://sraf.nd.edu>

The Use of Word Lists in Textual Analysis, Bill McDonald with Tim Loughran, *Journal of Behavioral Finance*, 16:1, Jan-Mar 2015, 1-11.

Textual Analysis in Accounting and Finance: A Survey, Bill McDonald with Tim Loughran, *Journal of Accounting Research*, 54:4, September 2016, 1187-1230.

Measuring Readability in Financial Disclosures, Bill McDonald with Tim Loughran, *Journal of Finance*, 69:4, August 2014, 1643-1671.

Regulation and Financial Disclosure: The Impact of Plain English, Bill McDonald with Tim Loughran, *Journal of Regulatory Economics*, 45:1, 2014, 94-113.

Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks, Bill McDonald with Priyank Gandhi and Tim Loughran, *Journal of Behavioral Finance*, forthcoming.

The Use of EDGAR Filings by Investors, Bill McDonald with Tim Loughran, *Journal of Behavioral Finance*, Vol. 18:2, 2017, 231-248.

Textual Analysis in Accounting and Finance: A Survey, Bill McDonald with Tim Loughran, *Journal of Accounting Research*, 54:4, September 2016, 1187-1230.

When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, Tim Loughran with Bill McDonald, *Journal of Finance*, vol. 66, February 2011, 35-65.

Using 10-K Text to Gauge Financial Constraints, Bill McDonald with Andriy Bodnaruk and Tim Loughran, *Journal of Financial and Quantitative Analysis*, 50:4, August. 2015, 1-24.

Trust, but Verify: MD&A Language and the Role of Trust in Corporate Culture, Bill McDonald with Tim Loughran and Robert Audi, *Journal of Business Ethics*, 18:2, 2017, 1187-1230.

A Wolf in Sheep's Clothing: The Use of Ethics-Related Terms in 10-K Reports, Tim Loughran with Bill McDonald and Hayong Yun, *Journal of Business Ethics*, vol. 89, 2009, 39-49.

Who Benefited from the Disclosure Mandates of the 1964 Securities Acts Amendments?, Tim Loughran with Robert Battalio and Brian Hatch, *Journal of Corporate Finance*, vol. 17, 2011, 1047-1063.

Helpful Natural Language Processing and Application Summaries and Surveys

Fisher, I.E., Garnsey, M.R., and Hughes, M.E. (2016). Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intelligent Systems in Accounting, Finance and Management*, Vol. 23, Issue 3, March 2016.

Kearney, C., Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, Vol. 33, May 2014, pp. 171-185.

Li, F (2010). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature*, Vol. 29, pp. 143 - 165.

Finance, Accounting, and Organizational Management:

Bonsall, S. B., Leone, A., Miller, B. P., & Rennekamp, K. (2017). A Plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3), 329-357

Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H., and Steele, L.B. (2014). The information content of mandatory risk factor disclosures in corporate filings, *Review of Accounting Studies*, Vol. 19, Issue 1, March 2014, pp. 396-455.

Dyer, T., Lang, M., and Lang, M. (2017) The evolution of 10-K disclosure: Evidence from Latent Dirichlet Allocation, *Journal of Accounting and Economics*, Vol. 64, Issue 2-3, November 2017, pp. 221-245.

Eccles, R.G., Krzus, M.K., Rogers, J., and Serafeim, G. (2012) The need for sector-specific Materiality and Sustainability reporting standards, *Journal of applied Corporate Finance*, Vol. 24, Issue 2, July 2012.

Frankel, Richard M. and Jennings, Jared N. and Lee, Joshua A., Using Natural Language Processing to Assess Text Usefulness to Readers: The Case of Conference Calls and Earnings Prediction (January 17, 2017).

Gerdes, J. Jr., (2003) EDGAR-Analyzer: automating the analysis of corporate data contained in the SEC's EDGAR database, *Decision Support Systems*, Vol. 35, Issue 1, April 2003, pp. 7-29. [*A helpful summary of automated analysis outside the Loughran and McDonald frameworks herein included*]

Goel, S. and Gangolly, J. (2012). Beyond the numbers: Mining the annual reports for hidden cues indicative of financial statement fraud, *Intelligent Systems in Accounting, Finance and Management*, Vol. 19, Issue 2, June 2012.

Jegadeesh, N. (2013) Word power: A new approach for content analysis, *Journal of Financial Economics*, Vol. 110, Issue 3, December 2013, pp. 712-729.

Lam, M. & Morriss, J. (2011). Do Two Negatives Make Good News or Worse News? Extending Textual Analysis of Corporate Disclosures beyond Counting Words. Academy of Financial Services Conference, Las Vegas, Nevada.(presentation)

Lam, M. & Morriss, J. (2011). Finding Meaning and Value in Corporate Disclosures. Proceedings of the Northeast Business & Economics Association, pp. 259-262. (presentation)

Li, F. (2008). Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics*, Vol. 45, Issue 2-3, August 2008, pp. 221-247.

Li, F. (2010). The information content of forward-looking statements in corporate filings--A Naive Bayesian machine learning algorithm approach, *Journal of Accounting Research* 48 (December 2010) pp. 1049-1102.

Nelson, Karen K. and Pritchard, Adam C., Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language (August 2007). 2nd Annual Conference on Empirical Legal Studies Paper.

Peters, G.F., and Romi, A.M. (2013) Discretionary compliance with mandatory environmental disclosures: Evidence from SEC filings, *Journal of Accounting and Public Policy*, Vol. 32, Issue 4, July-August 2013, pp. 213-236.

Richard A. Cazier and Ray J. Pfeiffer (2016) Why are 10-K Filings So Long?. *Accounting Horizons*: March 2016, Vol. 30, No. 1, pp. 1-21.

Starks, L.T. (2009). EFA Keynote Speech: "Corporate governance and corporate social responsibility What do investors care about? What should investors care about?" *Financial Review*, Vol. 44, Issue 4, October 2009.

Legal:

Citizens United v. Federal Election Commission, 558 U.S. 310 (more)130 S. Ct. 876; 175 L. Ed. 2d 753; 2010 U.S. LEXIS 766.

Backer, L.C. (2003). The Duty to Monitor: Emerging Obligations of Outside Lawyers and Auditors to Detect and Report Corporate Wrongdoing Beyond the Federal Securities Laws, 77 *St. John's L. Rev.* 919.

Bebchuk, L.A. and Jackson, R.J.Jr. (2010). Corporate Political Speech, Who Decides? *Harvard Law Review*, Vol. 124, No. 1, November 2010, pp. 83-117.

Humes, A. (2006-2007). SEC v. Siebel Systems, Inc.: Comment on the SEC's Failed Enforcement of Regulation Fair Disclosure in Federal Court, 32 *J. Corp. L.* 161.

Melone, M.A. (2010-2011). Citizens United and Corporate Political Speech: Did the Supreme Court Enhance Political Discourse or Invite Corruption?, 60 *DePaul L. Rev.* 29.

Redish, M.H. and Wasserman, H.M. (1997-1998). What's Good for General Motors: Corporate Speech and the Theory of Free Expression, 66 Geo. Wash. L. Rev. 235.

Siebecker, M.R. (2006-2007). Corporate Speech, Securities Regulation, and an Institutional Approach to the First Amendment, 48 Wm. & Mary L. Rev. 613.

Tucker, A. (2010-2011). Flawed Assumptions: A Corporate Law Analysis of Free Speech and Corporate Personhood in Citizens United, 61 Case W. Res. L. Rev. 497.

Winkler, A. (2006-2007). Corporate Personhood and the Rights of Corporate Speech, 30 Seattle U. L. Rev. 863.

Yang, K. and Page, A. (2005-2006). Controlling Corporate SPeech: Is Regulation Fair Disclosure Unconstitutional?, 39 U.C. Davis L. Rev. 1.

A common mistake that people make when trying to design something completely foolproof is to underestimate the ingenuity of complete fools.
Douglas Adams

APPENDICES

Due to the length and form of appended supplements, each has been attached with cover page and description following the final page of this document.

Appendix 1: **NLP Code**

PDF Printout of the Natural Language Processing Python Code used to pull the words as data for this project. Includes: initial package calls and downloads, paragraph by paragraph counting mechanism, regulatory word list counting and vector input code, “Blame” word counting mechanism, and Out File CSV closure program.

Appendix 2: **Sample Out File**

Columnized CSV file output, shows column by column sample data produced from a single year reporting pull from the NLP Code.

Appendix 3: **SEC SIC Firm Groups List**

While a full list of SEC SIC Codes can be found on the “Division of Corporation Finance: Standard Industrial Classification (SIC) Code List” at: www.sec.gov/info/edgar/siccodes.htm, this appendix includes the subset of SIC codes used in the 3-industry comparison employed in this project.

Appendix 4: **3 Group SIC Firm Group Comparison Full Data Table**

Full table of the data compiled in this project for the 3-industry comparison employed in this project. Also includes a comparison of the “highest overall” from the entire sample, with firm name and relevant SIC, at right for reference.

He gazed keenly into the distance and looked as if he would quite like the wind to blow his hair back dramatically at that point, but the wind was busy fooling around with some leaves a little way off.
Douglas Adams

-- END --

I may not have gone where I intended to go, but I think I have ended up where I needed to be.
Douglas Adams

Appendix 1 - NLP Code PDF Cover Page

Python Code pdf follows.

*Isn't it enough to see that a garden is beautiful without having to believe that there are fairies at the bottom
of it too?*

Douglas Adam


```

# -*- coding: utf-8 -*-
"""
Spyder Editor

This is a temporary script file.
"""

#In Core Packages

#Overall Progs
import numpy
import os.path
import urllib.request
from bs4 import BeautifulSoup, SoupStrainer
import re
import zipfile
import sys
import webbrowser
import urllib.request
import codecs
import csv
import sys
import collections
import glob
import string
from string import punctuation

#Word List Progs
import os
import shutil
import nltk
from string import punctuation
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import stopwords
from nltk.tokenize import RegexpTokenizer
from collections import OrderedDict
from sklearn.feature_extraction.text import TfidfVectorizer
from collections import defaultdict
import re

#Set two word functions for help later with NLTK
tokenizer= RegexpTokenizer(r'\w+') #Splits up strings into little strings
NOPunct = set(string.punctuation) #set of punctuation symbols for later removal

#####
#Step 1: Load Word Vectors from word Lists #####
#####

#Open Empty Vectors for Word Lists (note: NOT STEMMING since reg titles full)
RegWords= [] #Storage for General Regulation words list
DueToWords = [] #Storage for Blame words list

#Path to Word List.csv files
Wordpath = "/Users/KMills/Desktop/Hart Fellowship FALL/"

```

```

#Word File Names
RegCSV = "LawWords_Hart2018.csv"
DueToCSV = "DueToWords_Hart2018.csv"

#Setting General Reg Words Set into RegWords vector

with open(Wordpath+RegCSV,"r") as reg:
    line=reg.readline().rstrip('\n')
    RegWords.append(line) #Adding each word to our word list for positive words
    while(line):
        if line !=RegWords[-1]: #Checking for repeats before we place things into
            the list
            RegWords.append(line) #adds the new words only when it is not a repeat
            line=reg.readline().rstrip('\n')

        #Testing the the List of regulatory words

        print(line)
    print(len(RegWords)) #Notes the number of pos. words (unstemmed)
    reg.close()

#Setting Blame (DueTo) Words Set into DueToWords vector

with open(Wordpath+DueToCSV,"r") as dt:
    line=dt.readline().rstrip('\n')
    DueToWords.append(line) #Adding each word to our word list for positive words
    while(line):
        if line !=DueToWords[-1]: #Checking for repeats before we place things into
            the list
            DueToWords.append(line) #adds the new words only when it is not a repeat
            line=dt.readline().rstrip('\n')

        #Testing the the List of regulatory words

        print(line)
    print(len(DueToWords)) #Notes the number of pos. words (unstemmed)
    reg.close()

print("Word lists converted into Vector Form")

#####
##Step 2: Send the program to go return data #####
#####

#Setting path to 10-K Files
PracticePath = "/Users/KMills/Desktop/Sample_Processed"

# Setting matrix to house each data line (document by document)
StatLines = []

#Sending the program into the files in loop
for file in os.listdir(PracticePath):          # Selects the file from the
    directory file

```

```

if file.endswith(".txt"):          # Opens each one of the txt files in the
    directory
    #print("CURRENT FILE: "+os.path.join(PracticePath,file)) # Prints the
    #path to the current FILE
    doc = open(os.path.join(PracticePath,file)) #Opens the individual 10K
    #file itself

    print("File Opened: "+os.path.join(PracticePath,file)) #Initial reg File
    #Opened Sanity Check

    #####
    #### Setting the initial empty vector for word string for file ####
    #####

    DocCorpus = [] #setting clean repository for reg-word-wholetext
    print("-- [ ] -- Empty Corpus Opened")
    MiniCorpus = [] #opening initial mini-corpus
    #will have writing procesude over mini-corpus later

    #####
    ### Words_Flagd: Reg or Act words counted          ###
    ### Blame_Words: Blame words counted in reg/act paragraphs ###
    ### Para_Count: Total Paragraphs Counter          ###
    ### RegPara_Count: Total Paragraphs with Reg/Act Word    ###
    #####

    #Setting Word Counters to Zero
    Words_Flagd = 0
    Blame_Words = 0
    Para_Count = 0
    RegPara_Count = 0

    #Setting word list (General Reg or Act Words) to be used
    List_Name = RegWords #use same code for list of gen/specific

    #####
    #### Sending the program in to read lines in the cleaned form ####
    #####

    lines = doc.readlines()    #sets lines as reading line-by-line
    print ("Quiet! I'm Trying to Read these lines!!") #Read lines sanity
    #check

    for line in lines:

        #####
        #### Grabbing Vitals before beginning the lanague processing ####
        #####

        #####
        ### CIK Number coded to: CIK_Number          ###
        ### Report Date coded to: Rep_Date            ###
        #####

```

```

### Company Name coded to: Co_Name          #####
### SIC Number coded to: SIC_Number         #####
#####

if "ACCESSION NUMBER:" in line and re.findall(r"D(\d{10})D",line):

    CIK_Number = str(re.findall(r"D(\d{10})D",line))
    print("CIK Located: "+CIK_Number)

if "FILED AS OF DATE:" in line and re.findall(r"D(\d{8})D",line):

    Rep_Date = str(re.findall(r"D(\d{8})D",line))
    print("Report Date Located: "+Rep_Date)

if "COMPANY CONFORMED NAME:" in line:

    try:

        Name_Line=line.split("\t\t\t")
        #print(str(Name_Line)) #Line structure
        #print(str(Name_Line[1])) #Name structure
        Dirty_Name = Name_Line[1] #Un-Chomped Name
        Co_Name = str(Dirty_Name).rstrip() #Chomps new line char
        print("Company Name Found: "+Co_Name) #Update Check

    except:
        Co_Name = "Null_Split_Co_Name"
        print(Co_Name)

if "STANDARD INDUSTRIAL CLASSIFICATION:" in line and
re.findall(r"D(\d{4})D",line):

    SIC_Number = str(re.findall(r"D(\d{4})D",line))
    print("SIK Found: "+SIC_Number)

#####
## Now setting up the word counting process for internal corpus ##
#####

#####
### Number of Regulatory/Act Words: Words_Flagd      #####
### Number of Blame Words in text: Blame_Words       #####
### Length of Flagged Text string: Str_Leng          #####
#####

if re.search('[a-zA-Z]', line):

    Words = line.rstrip()
    MiniCorpus.append(Words)

    print(MiniCorpus)

else:

```

```

#####
## If we hit a new paragraph, we want to:      ###
## 1. Test the Mini-Corpus for Reg Words      ###
## 2. If we find a reg word we append to Doc Corpus ###
## 3. If we find NO Reg Word, we move on to new para ###
## -- So each Mini-Corpus is going to be appended ###
## to the main corpus ONLY IF it contains a reg word ###
## -- Counts the number of paragraphs total (ie total ###
## tested paragraphs) and then counts the number of the ###
## paragraphs we have tested which also contain reg/act ###
## words - this gives us a good look at volume of reg ###
## speech relative to document length for each txt file ###
#####

print("Blank Space Found -- TESTING PARAGRAPH")

Para_Count = Para_Count + 1

Dirty_Words = str(MiniCorpus)
words = ''.join(c for c in Dirty_Words if c not in NOPunct)

for word in List_Name:

    WC = words.count(word) #Counts the offuracne of the word

    #print(str(word)+": "+str(WC)) #tracking as we pull words

    Words_Flagd = Words_Flagd + WC #Updates overall counter

    if WC > 0:

        DocCorpus.append(MiniCorpus)

        #DocCorpusTracker = DocCorpusTracker + 1

        #print("Document Corpus Updated: "+str(DocCorpus))

        RegPara_Count = RegPara_Count + 1

MiniCorpus = []

print("MC Reset for New")

#####
### Now we want to text the overall corpus of regulatory #
### paragraphs for the bame words. So now we want to take our #
### DocCorpus and do a word count inside that for the words #
### that we have labeled as blame words :- ) #
#####

Dirty_Corpus = str(DocCorpus)

corpus = ''.join(c for c in Dirty_Corpus if c not in NOPunct )

```

```

Str_Leng = len(corpus)

for blameword in DueToWords:

    BC = corpus.count(blameword)

    #print(str(blameword)+" : "+str(BC)) #tracking as we pull words

    Blame_Words = Blame_Words + BC #Updates overall counter

#####
### Now creating tie line which will be added to our matrix for ##
### the data and counts associated with this document:          ##
### Entity and Report Data, Word Counts, and Lengths, etc.      ##
#####

#Cleaning a few of my strings before we append to csv file

Report_Date = str(Rep_Date)
Report_Date = ''.join(c for c in Report_Date if c not in NOPunct )

Company_Name = str(Co_Name)
Company_Name = ''.join(c for c in Company_Name if c not in NOPunct )

CIKNumber = str(CIK_Number)
CIKNumber = ''.join(c for c in CIKNumber if c not in NOPunct )

SICNumber = str(SIC_Number)
SICNumber = ''.join(c for c in SICNumber if c not in NOPunct )

#Creating call line Fact list:
ID_Line = str(Report_Date)+","+str(Company_Name)+","+str(CIKNumber)
          +","+str(SICNumber)
Data_Line = str(Para_Count)+","+str(RegPara_Count)+","+str(Words_Flagd)
          +","+str(Str_Leng)+","+str(Blame_Words)
Stats_Line = str(ID_Line+","+Data_Line)

print(Stats_Line)

#####
### Appending the stat-line to the stat line Matrix for overall ###
#####

StatLines.append(Stats_Line)

print("Stats Line Appended total files: "+str(len(Stats_Line)))

#####
##### Closing doc before heading on to next pull file #####
#####

doc.close
print("File: "+os.path.join(PracticePath,file)+" CLOSED")

```

```
#####
### Now printing the StatLines to a CSV file so we can actually access the ###
### data that we have created :-)
```

#####

```
#Creating the output CSV file
LocalOut = "/Users/KMills/Desktop/SAMPLEOUT/" #Local Output file folder
Stats_CSV = "10K_SampleOut_Hart2018.csv"

GoTo = str(str(LocalOut)+str(Stats_CSV)) #Directions for outfile

with open(GoTo, "w") as OutCSV:
    wr = csv.writer(OutCSV, lineterminator='\n')
    for stat in StatLines:
        wr.writerow([stat])

    OutCSV.close()
```

Appendix 2 - Sample Out-File CSV Table

Report Date	Firm Name	CIK Number	SIC Number	Total Paragraphs	Regulatory Paragraphs	# Reg Words	Reg String Length	# Blame Words
20080331	ADAMS RESOURCES ENERGY INC	2178	5172	3237	10	10	810	1417
20080228	ALEXANDER BALDWIN INC	3453	4400	47004	13	15	1589	44021
20080225	ALEXANDERS INC	3499	6798	6563	20	21	5523	9039
20080604	AMERCO NV	4457	7510	4614	3	3	274	0
20080220	AMR CORP	4515	4512	8718	4	4	243	0
20080220	AMERICAN AIRLINES INC	4515	4512	7475	4	4	243	0
20080228	AMERICAN ELECTRIC POWER CO INC	4904	4911	19824	38	39	5126	28361
20080228	OHIO POWER CO	4904	4911	19732	35	36	5089	28588
20080228	APPALACHIAN POWER CO	4904	4911	19922	38	39	5130	28509
20080228	PUBLIC SERVICE CO OF OKLAHOMA	4904	4911	19139	38	39	5220	28555
20080228	COLUMBUS SOUTHERN POWER CO OF	4904	4911	19437	40	43	7508	343887
20080228	SOUTHWESTERN ELECTRIC POWER CO	4904	4911	19109	37	38	5283	28553
20080228	INDIANA MICHIGAN POWER CO	4904	4911	19080	37	38	5121	28227
20080328	AEP Texas Central Transition Funding I	4904	6189	566	5	5	460	0
20080714	AMREP CORP	6207	6500	1146	11	11	8709	2549
20080305	ANALYSTS INTERNATIONAL CORP	6292	7371	3108	5	5	455	835
20080401	STAGE STORES INC	6885	5651	4041	17	17	2066	0
20080314	ARABIAN AMERICAN DEVELOPMENT C	7039	2911	3390	12	12	1126	209
20080415	ARABIAN AMERICAN DEVELOPMENT C	7039	2911	434	5	5	417	603
20080829	ARCHER DANIELS MIDLAND CO	7084	2070	4464	12	12	663	0
20080228	SOUTHWESTERN ENERGY CO	7332	1311	17816	12	16	9077	5274
20081126	ATWOOD OCEANICS INC	8411	1381	7116	29	30	35266	75362
20080512	AZZ INC	8947	3640	3179	6	6	590	179
20080404	BAGDAD CHASE INC	9128	1000	240	4	4	244	0
20081118	BAGDAD CHASE INC	9128	1000	243	4	4	245	0
20080314	BALDWIN LYONS INC	9346	6331	35697	8	8	381	931

*It is a mistake to think you can solve any major problems just with potatoes.
Douglas Adams*

Appendix 3 - SEC SIC Codes List Firm Groups Selected for Comparison

SIC Subsets Chosen:

1311 - 1389: Petroleum and Natural Gas

5000 - 5990: Wholesale and Retail

6021 - 6282: Financial and Investment Institutions

1311-1389 Petroleum and Natural Gas

[All Overseen by A/D Office 4]

1311: CRUDE PETROLEUM AND NATURAL GAS

1381: DRILLING OIL AND GAS WELLS

1382: OIL AND GAS FIELD EXPLORATION

1389: OIL AND GAS FIELD SERVICES, NEC

5000-5990: Wholesale and Retail

[All Overseen by A/D Offices: 2, 3, 4, 5, 6, 9, 10]

5000: WHOLESALE-DURABLE GOODS

5010: WHOLESALE-MOTOR VEHICLES & MOTOR VEHICLE PARTS &

SUPPLIES

5013: WHOLESALE-MOTOR VEHICLE SUPPLIES & NEW PARTS

5020: WHOLESALE-FURNITURE & HOME FURNISHINGS

5030: WHOLESALE-LUMBER & OTHER CONSTRUCTION MATERIALS

5031: WHOLESALE-LUMBER, PLYWOOD, MILLWORK & WOOD PANELS

5040: WHOLESALE-PROFESSIONAL & COMMERCIAL EQUIPMENT &

SUPPLIES

5045: WHOLESALE-COMPUTERS & PERIPHERAL EQUIPMENT & SOFTWARE

5047: WHOLESALE-MEDICAL, DENTAL & HOSPITAL EQUIPMENT & SUPPLIES

5050: WHOLESALE-METALS & MINERALS (NO PETROLEUM)

5051: WHOLESALE-METALS SERVICE CENTERS & OFFICES

5063: WHOLESALE-ELECTRICAL APPARATUS & EQUIPMENT, WIRING

SUPPLIES

5064: WHOLESALE-ELECTRICAL APPLIANCES, TV & RADIO SETS

5065: WHOLESALE-ELECTRONIC PARTS & EQUIPMENT, NEC

5070: WHOLESALE-HARDWARE & PLUMBING & HEATING EQUIPMENT &
SUPPLIES

5072: WHOLESALE-HARDWARE

5080: WHOLESALE-MACHINERY, EQUIPMENT & SUPPLIES

5082: WHOLESALE-CONSTRUCTION & MINING (NO PETRO) MACHINERY &
EQUIP

5084: WHOLESALE-INDUSTRIAL MACHINERY & EQUIPMENT

5090: WHOLESALE-MISC DURABLE GOODS

5094: WHOLESALE-JEWELRY, WATCHES, PRECIOUS STONES & METALS

5099: WHOLESALE-DURABLE GOODS, NEC

5110: WHOLESALE-PAPER & PAPER PRODUCTS
 5122: WHOLESALE-DRUGS, PROPRIETARIES & DRUGGISTS' SUNDRIES
 5130: WHOLESALE-APPAREL, PIECE GOODS & NOTIONS
 5140: WHOLESALE-GROCERIES & RELATED PRODUCTS
 5141: WHOLESALE-GROCERIES, GENERAL LINE
 5150: WHOLESALE-FARM PRODUCT RAW MATERIALS
 5160: WHOLESALE-CHEMICALS & ALLIED PRODUCTS
 5171: WHOLESALE-PETROLEUM BULK STATIONS & TERMINALS
 5172: WHOLESALE-PETROLEUM & PETROLEUM PRODUCTS (NO BULK STATIONS)
 5180: WHOLESALE-BEER, WINE & DISTILLED ALCOHOLIC BEVERAGES
 5190: WHOLESALE-MISCELLANEOUS NONDURABLE GOODS
 5200: RETAIL-BUILDING MATERIALS, HARDWARE, GARDEN SUPPLY
 5211: RETAIL-LUMBER & OTHER BUILDING MATERIALS DEALERS
 5271: RETAIL-MOBILE HOME DEALERS
 5311: RETAIL-DEPARTMENT STORES
 5331: RETAIL-VARIETY STORES
 5399: RETAIL-MISC GENERAL MERCHANDISE STORES
 5400: RETAIL-FOOD STORES
 5411: RETAIL-GROCERY STORES
 5412: RETAIL-CONVENIENCE STORES
 5500: RETAIL-AUTO DEALERS & GASOLINE STATIONS
 5531: RETAIL-AUTO & HOME SUPPLY STORES
 5600: RETAIL-APPAREL & ACCESSORY STORES
 5621: RETAIL-WOMEN'S CLOTHING STORES
 5651: RETAIL-FAMILY CLOTHING STORES
 5661: RETAIL-SHOE STORES
 5700: RETAIL-HOME FURNITURE, FURNISHINGS & EQUIPMENT STORES
 5712: RETAIL-FURNITURE STORES
 5731: RETAIL-RADIO, TV & CONSUMER ELECTRONICS STORES
 5734: RETAIL-COMPUTER & COMPUTER SOFTWARE STORES
 5735: RETAIL-RECORD & PRERECORDED TAPE STORES
 5810: RETAIL-EATING & DRINKING PLACES
 5812: RETAIL-EATING PLACES
 5900: RETAIL-MISCELLANEOUS RETAIL
 5912: RETAIL-DRUG STORES AND PROPRIETARY STORES
 5940: RETAIL-MISCELLANEOUS SHOPPING GOODS STORES
 5944: RETAIL-JEWELRY STORES
 5945: RETAIL-HOBBY, TOY & GAME SHOPS
 5960: RETAIL-NONSTORE RETAILERS
 5961: RETAIL-CATALOG & MAIL-ORDER HOUSES
 5990: RETAIL-RETAIL STORES, NEC

6021 - 6282: Financial and Investment Institutions

[All Overseen by A/D Offices: 7, 8, and OSF (Office of Structured Finance)]

6021: NATIONAL COMMERCIAL BANKS
6022: STATE COMMERCIAL BANKS
6029: COMMERCIAL BANKS, NEC
6035: SAVINGS INSTITUTION, FEDERALLY CHARTERED
6036: SAVINGS INSTITUTIONS, NOT FEDERALLY CHARTERED
6099: FUNCTIONS RELATED TO DEPOSITORY BANKING, NEC
6111: FEDERAL & FEDERALLY-SPONSORED CREDIT AGENCIES
6141: PERSONAL CREDIT INSTITUTIONS
6153: SHORT-TERM BUSINESS CREDIT INSTITUTIONS
6159: MISCELLANEOUS BUSINESS CREDIT INSTITUTION
6162: MORTGAGE BANKERS & LOAN CORRESPONDENTS
6163: LOAN BROKERS
6172: FINANCE LESSORS
6189: ASSET-BACKED SECURITIES
6199: FINANCE SERVICES
6200: SECURITY & COMMODITY BROKERS, DEALERS, EXCHANGES &
SERVICES
6211: SECURITY BROKERS, DEALERS & FLOTATION COMPANIES
6221: COMMODITY CONTRACTS BROKERS & DEALERS
6282: INVESTMENT ADVICE

Shee, you guys are so unhip it's a wonder your bums don't fall off.
Douglas Adams

Appendix 4 - 3 Group SIC Firm Group Comparison Full Data Table Cover Page
Full Data Table pdf follows.

Those who study the complex interplay of cause and effect in the history of the Universe say that this sort of thing is going on all the time, but that we are powerless to prevent it. 'It's just life,' they say.

Douglas Adams

SIC: 1311 - 1389 Petroleum / Natural Gas								SIC: 5000 - 5990 Wholesale and Retail								SIC: 6021 - 6282 Financial and Investment Institutions								All Filings	
Year	Total Filings	Average Paragraphs	Average Reg. Paragraphs	Average Number Reg Words Flag'd	Average Reg Paragraphs Total Words	Average "Blame" words	Reg Paragraphs MAX (Blame Words Rec'd)	Total Filings	Average Paragraphs	Average Reg. Paragraphs	Average Number Reg Words Flag'd	Average Reg Paragraphs Total Words	Average "Blame" words	Reg Paragraphs MAX (Blame Words Rec'd)	Total Filings	Average Paragraphs	Average Reg. Paragraphs	Average Number Reg Words Flag'd	Average Reg Paragraphs Total Words	Average "Blame" words	Reg Paragraphs MAX (Blame Words Rec'd)	Reg Paragraphs MAX (Blame Words Rec'd)	Max Entity Notes		
1997	737	779.785617	10.9213026	12.421981	2960.10312	12562.3541	139 (607742 Blame Words)	1555	1000.6	12.32797428	14.268881029	3511.516399	15210.58585	196 (1910191 Blame Words)	2418	942.0248139	11.15301902	12.11538462	2782.194376	13452.55211	415 (2704135 Blame Words)	581 (135699 Blame Words)	ADVANCED MICRO DEVICES INC.SIC: 3674		
1998	631	863.797147	12.5419968	14.4405705	3269.35499	12102.9271	277 (1750752 Blame Words)	1553	996.998712	12.6735351	14.8126207	3639.50547	18949.8609	182 (2072706 Blame Words)	2519	934.660183	11.2572449	12.3068678	3157.70226	14651.9766	234 (8063379 Blame Words)	422 (194164 Blame Words)	ACTV INC DE SIC: 3663		
1999	656	929.797256	12.5381098	14.6935976	3339.82317	13775.2012	187 (155433 Blame Words)	1421	1006.70795	12.9711471	15.5003519	4146.85292	14926.6341	345 (1183389 Blame Words)	2677	846.433321	10.0048562	10.8528203	2329.74785	7306.22899	166 (72778 Blame Words)	576 (88304 Blame Words)	COLTIER PHARMACEUTICALS INC. SIC: 2834		
2000	643	904.503888	11.6982893	13.5381026	3631.45879	9711.42457	310 (54653 Blame Words)	1505	1024.85847	12.3548173	14.5169435	3759.65249	15772.6246	210 (1948555 Blame Words)	2417	944.352089	10.4038064	11.3491932	2683.46421	9252.07944	192 (353662 Blame Words)	357 (2364311 Blame Words)	HUNT CORP SIC: 3950		
2001	586	1026.38567	11.6843003	13.5443686	4216.72014	10502.7218	130 (117548 Blame Words)	1402	1212.74822	11.6718973	13.7154066	4971.05777	10438.6805	159 (381873 Blame Words)	2294	1203.05013	10.097646	11.005667	3003.96295	9745.77158	125 (935165 Blame Words)	990 (35354325 Blame Words)	AMTRAN INC.SIC:4522		
2002	591	1369.70728	11.8680203	13.7275804	4437.49239	13085.824	227 (1102305 Blame Words)	1334	2201.91004	11.8223388	13.7151424	4241.24588	12918.3943	163 (470347 Blame Words)	2761	1790.66135	8.93227092	10.152481	2668.74973	10948.2945	152 (284865 Blame Words)	539 (7539106 Blame Words)	CALPINE CORP SIC: 4911		
2003	539	2293.33395	12.8534323	14.8089054	4483.47124	15530.0445	89 (174063 Blame Words)	1250	3197.664	14.0256	16.3488	5875.5912	22309.8264	202 (315462 Blame Words)	2963	2807.06682	9.68275397	11.0283496	3980.2511	17372.4533	338 (7075870 Blame Words)	740 (815844 Blame Words)	CREE INC.SIC: 3674		
2004	580	3251.18276	14.0344828	16.2965517	5898.55172	22286.6638	330 (3825094 Blame Words)	1172	3955.75	13.5827645	16.1032423	5634.85666	12953.1911	169 (574601 Blame Words)	3441	3259.1764	9.22667829	10.5655333	4084.04999	20260.2517	132 (30192 Blame Words)	534 (5820482 Blame Words)	INLAND WESTERN RETAIL REAL ESTATE TRUST INC.SIC:6798		
2005	651	4264.23502	11.2688172	12.59447	4280.80952	10916.4501	87 (268326 Blame Words)	1272	4636.3066	11.3254717	12.7727987	4890.32626	12012.4245	158 (444378 Blame Words)	3751	3673.21434	8.54118902	9.73873634	3524.19328	17011.851	150 (4751642 Blame Words)	988 (16859 Blame Words)	MOJUNA HEALTHCARE INC.SIC:6324		
2006	671	5272.90015	10.7108793	12.5424739	4317.92846	9706.50671	201 (94556 Blame Words)	1163	6002.09286	11.1117799	12.4075666	4637.27515	13125.0542	193 (6606974 Blame Words)	3656	4207.02899	8.2095186	9.03637856	3204.77188	9812.42287	259 (866950 Blame Words)	348 (0 Blame Words)	TRAMMELL CROW CO SIC:6510		
2007	656	6568.45579	10.5960366	11.7439024	4557.0122	17609.8643	119 (5108839 Blame Words)	1071	7178.53595	11.2418301	12.3043884	5521.87488	7098.9141	444 (346794 Blame Words)	3404	5636.50558	8.5126322	9.56169213	4233.90159	19050.8305	193 (3294244 Blame Words)	574 (159756 Blame Words)	LEXINGTON REALTY TRUST SIC:6798		
2008	718	6726.92479	9.91643454	10.9791086	4395.90808	8911.58357	137 (846651 Blame Words)	1136	6543.80986	10.0114437	10.909331	3903.97447	9742.48063	98 (47232 Blame Words)	2936	6734.21764	8.22956403	9.02622616	3466.69074	16481.8634	129 (126724 Blame Words)	596 (197278 Blame Words)	CERUS CORP.SIC:3841		
2009	726	6587.66667	9.97520661	11.2644628	5417.46556	6469.07163	280 (67302 Blame Words)	999	7224.15015	10.1091091	11.2082082	4376.33634	10931.1892	90 (113721 Blame Words)	1842	10687.221	10.0570033	11.0206298	5260.02606	28869.9799	145 (24383804 Blame Words)	318 (1234890 Blame Words)	PHARMATHENE INC. SIC:2834		
2010	691	6344.13893	9.17945007	10.3545586	12494.7236	8432.50507	158 (564 Blame Words)	985	7408.33604	10.3472081	11.5015228	12071.6985	18524.7919	152 (4032216 Blame Words)	1603	11706.1085	10.0692452	10.9201497	8854.53712	18517.0618	146 (603562 Blame Words)	668 (174887 Blame Words)	LOUISIANAPACIFIC CORP SIC:2400		
2011	620	8778.11774	9.98402556	11.3051118	16387.976	13447.4473	123 (1675569 Blame Words)	734	9110.490463	10.98228883	12.35694823	13530.24251	19401.6921	96 (29732 Blame Words)	1482	15009.9825	10.3306343	11.2253711	21020.5891	41722.3846	107 (158863 Blame Words)	485 (2559536 Blame Words)	GENERAL GROWTH PROPERTIES INC.SIC: 6798		
2012	578	8599.9654	12.866782	14.8633218	66976.2872	36222.9291	85 (558858 Blame Words)	886	9059.49323	13.1918736	14.6839729	59978.1095	36364.7302	146 (0 blame words)	1490	14514.4705	13.452349	15.3785235	115192.335	72376.2993	271 (6324786 blame words)	570 (31665 Blame Words)	FORCE PROTECTION INC.SIC: 3790		