

Text analysis on SEC filings (A course proposal)

Yuxing Yan (6/13/2016) ¹

Abstract

In this short paper, I propose a graduate course titled “Text Analysis on SEC filings”. There are several reasons why we should offer such a course. First, the unstructured information² has a lion share of all information, 70% to 80% and it is reported that 80% of structured information came from unstructured one. Second, SEC filings is an important source of information (gold mine) since public companies, corporate insiders, and broker-dealers are required to make regular SEC filings.³ Third, from SEC filings we could retrieve both structured information, such as annual sales and net income, and unstructured information such as MD&A (Management Discussion and Comments). Fourth, SEC filings could be downloaded free of charge. Fifth, the tools used in this course are Perl and R, both of them are free as well. Sixth, via replication of several seminal papers, students would know how hard it is to collect and process the unstructured information and appreciate the benefits generated from unstructured data.

Structured information vs. unstructured information

Usually, we could classify information (data) into two categories: structured and unstructured. For structured data, we have a long history to use them. For finance and accounting, CREP and Compustat databases are two typical examples. CRSP stands for Center for Research in Security Prices. The database is generated and maintained by University Chicago. The database offers daily, monthly and annual information, such as price, return, trading volume and number of share outstanding, about the all exchanges listed American stocks from 1926 onwards. For example, from the following image, we know that CRSP database has data for a company called “Optimum Manufacturing Inc.) from 1/31/1986 to 6/30/1987.

```
> load("c:/temp/crspInfo.RData")
> head(crspInfo)
  PERMNO PERMCO  CUSIP          FIRMNAME TICKER EXCHANGE  BEGDATE  ENDDATE
1  10000   7952 68391610 OPTIMUM MANUFACTURING INC  OMFGA      3 1986-01-31 1987-06-30
2  10001   7953 36720410 GAS NATURAL INC          EGAS      2 1986-01-31 2011-12-30
3  10002   7954 05978R10 BANCTRUST FINANCIAL GROUP INC  BTFG      3 1986-01-31 2011-12-30
4  10003   7957 39031810 GREAT COUNTRY BK ASONIA CT    GCBK      3 1986-01-31 1995-12-29
5  10005   7961 95815510 WESTERN ENERGY RESOURCES INC  WERC      3 1986-01-31 1991-07-31
6  10006  22156 00080010 A C F INDUSTRIES INC          ACF       1 1925-12-31 1984-06-29
> |
```

Below is a typical example of non-quantitative (number) information.

Relative percentages of those two types of information

¹ Department of Economics and Finance, Canisius College, yany@canisius.edu

² https://en.wikipedia.org/wiki/Unstructured_data

³ <https://www.sec.gov/edgar.shtml>

Unstructured data is much more than structured data, 80% vs. 20% according to some sources.⁴ Computer World states that unstructured information might account for more than 70%–80% of all data in organizations.

Text analysis for finance and accounting

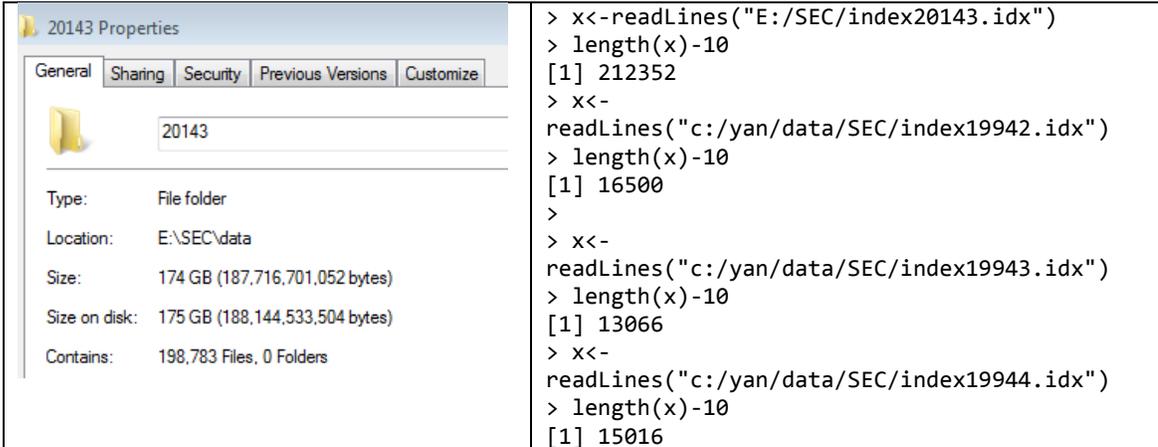
Applying text analysis to finance and accounting does not have a long history. Li (2008) shows that the readability of 10-K filings has a statistically significant impact on the performance of a firm's subsequent performance. Because of defining and measuring readability in the context of financial disclosures becomes important with the increasing use of textual analysis and the SEC's plain English initiative, Lougran and McDonald (2015) show that the Fog Index—the most commonly applied readability measure—is poorly specified in financial applications. Of Fog's two components, one is misspecified and the other is difficult to measure. They suggest to use the size of 10-K filing as a simple readability proxy and show that it outperforms the Fog Index. Another added advantage is that it does not require document parsing, thus facilitates replication. According to Loughran and McDonald (2015), there are 632 different forms.⁵ ON the other hand, most researchers used only one or two forms, such as 10-K. Thus, the SEC filings “database” is a gold mine waiting to be explored.

SEC filings

SEC stands for the U.S. Securities and Exchange Commission. According to law, public companies, certain insiders, and broker-dealers are required to make regular SEC filings, such as quarterly and annual financial statements or other formal documents.

Size of SEC filings

The size of SEC filings is huge. The size of SEC filings is quite big. For example the size of the quarterly index data is 218M while the size of just one quarter (2014 Q3) is 175G.



The image shows a Windows File Explorer window for a folder named '20143'. The folder is located at 'E:\SEC\data' and has a size of 174 GB (187,716,701,052 bytes). It contains 198,783 files and 0 folders. To the right of the File Explorer is a terminal window with the following R code and output:

```
> x<-readLines("E:/SEC/index20143.idx")
> length(x)-10
[1] 212352
> x<-
readLines("c:/yan/data/SEC/index19942.idx")
> length(x)-10
[1] 16500
>
> x<-
readLines("c:/yan/data/SEC/index19943.idx")
> length(x)-10
[1] 13066
> x<-
readLines("c:/yan/data/SEC/index19944.idx")
> length(x)-10
[1] 15016
```

From 1993 to 2016, we have 24 years, i.e., 94 quarters. If taking the average of quarterly in 1994Q2 (58G) and one quarterly data in 2014 (175G for Q3), as our one quarterly size, the total size is about 11T.

⁴ <https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>

⁵ http://www3.nd.edu/~mcdonald/Word_Lists.html or http://www3.nd.edu/~mcdonald/Word_Lists_files/LoughranMcDonald_MasterDictionary_2014.xlsx

Tools (languages) used for this course

In this course, we adopt both Perl and R as our working languages. Perl stands for Practical Extracting and Reporting Language which is a powerful scripting language for working with unformatted text. For the following reasons we have adopted those two languages. First, both Perl and R are free. Second, Perl is a perfect choice since its super ability to conduct a text analysis. R is chosen because its popularity in financial industry. Third, many students might be familiar with R already.

R program: <http://r-project.org>

Perl program: <http://www.activestate.com/>

DzSoft Perl editor: <http://dzsoft.com>

Target students

The ideal students are from finance, accounting majors. They should have taken at least one finance or accounting courses. In addition, they should understand R basics (or Perl basics) since two languages will be used for this course.

Prerequisites

- a) Business Analytics using R (FIN456A/MBA674A offered at Canisius)
- b) At least one finance or accounting course such as (FIN311, FIN414, FIN312)

Data sources

SEC filings: <https://www.sec.gov/edgar.shtml>

Word list related to finance and accounting (positive, negative words):

http://www3.nd.edu/~mcdonald/Word_Lists.html

List of potential term papers

The following table shows a list of potential topics for a term project.

#	Name
1	Download and analyze all SEC quarterly indices from 1993 2016Q1 using R (or Perl)
2	Download all 10-K from SEC Edgar
3	Download all filings for one year, such as 2015
4	Parse 10-K for year 2015
5	parse 13-f
6	Parse all forms (3,4 and 5)
7	Replicate Li (2008)
8	Replicate Loughran McDonald (2015)
9	Replicate the word list related to finance/accounting Loughran McDonald (2011)

References

Demers, Elizabeth, and Clara Vega, 2008, Soft information in earnings announcements: News or noise? Working paper, INSEAD.

- Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, Working paper, Northwestern University.
- Garcia, Diego and Oyvind Norli, 2012, Crawling EDGAR, working paper, UNC at Chapel Hill and Norwegian School of Management.
- Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal, 2008, The incremental information content of tone change in management discussion and analysis, Working paper, INSEAD.
- Griffin, Paul, 2003, Got information? Investor response to Form 10-K and Form 10-Q EDGAR filings, *Review of Accounting Studies* 8, 433–460.
- Hanley, Kathleen Weiss, and Gerard Hoberg, 2010, The information content of IPO prospectuses, *Review of Financial Studies* 23, 2821–2864.
- Henry, Elaine, 2008, Are investors influenced by the way earnings press releases are written? *Journal of Business Communication* 45, 363–407.
- Holzinger, Andreas; Stocker, Christof; Ofner, Bernhard; Prohaska, Gottfried; Brabenetz, Alberto; Hofmann-Wellenhof, Rainer (2013).
- Combining HCI, Natural Language Processing, and Knowledge Discovery – Potential of IBM Content Analytics as an Assistive Technology in the Biomedical Field, In Holzinger, Andreas; Pasi, Gabriella.
- Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Lecture Notes in Computer Science. Springer. pp. 13–24.
- Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.
- Li, Feng, 2009, The determinants and information content of the forward-looking statements in corporate filings—a Naive Bayesian machine learning approach, Working paper, University of Michigan.
- Li, Feng, 2010, Textual Analysis of Corporate Disclosures: A Survey of the Literature, *Journal of Accounting Literature* 29, 143-165.
- Loughran, Tim, Bill McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance*,
- Loughran, Tim and Bill McDonald, Natural Language Processing and Textual Analysis in Finance and Accounting, FMA presentation 2012.

Loughran, Tim and Bill McDonald, 2015, Measuring Readability in Financial Disclosures, *Journal of Finance* (forthcoming), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1920411

Mayew, William J., and Mohan Venkatachalam, 2009, The power of voice: Managerial affective states and future firm performance, Working paper, Duke University.

Routledge, Bryan R., Stefano Sacchetto, and Noah A. Smith, 2013, Predicting Merger Targets and Acquirers from Text, working paper, Carnegie Mellon University

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Tetlock, Paul C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

You, H., and X. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information, *Review of Accounting Studies* 14 4: 559–586.

Appendix A: R program to download just one index

add later

Appendix B: R program to get all quarterly indices

add later

Appendix C: A Perl program to download all SEC quarterly index files

add later

Appendix D: A Perl program to download all text files for a given quarterly index file

add later

Appendix E: Perl program for readability measures: Fog, Flesch and Flesch-Kincaid indices

add later

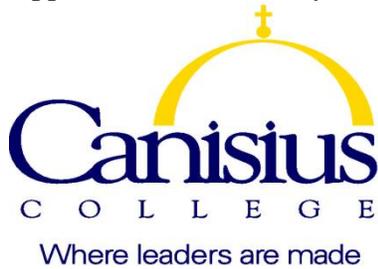
Appendix F: word frequency and word picture

```
> path<-  
'http://canisius.edu/~yany/doc/obama2008.txt'  
> x<-readTextFile(path)  
> y<-wordFrequency(x)  
> head(y)  
  black white   can many   one time  
      33   27   23   17   17   16  
> wordPicture(x)  
Loading required package: RColorBrewer
```

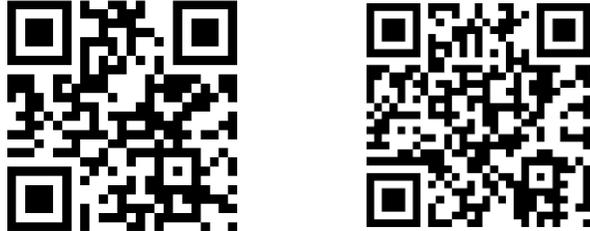


Appendix G: list of chapters

- chapter 1: Introduction to text analysis
- chapter 2: Introduction to SEC filings
- chapter 3: R basics
- chapter 4: Perl basics
- chapter 5: Regular expressions (Perl)
- chapter 6: Regular expressions (R)
- chapter 7: download and analyze SEC quarterly index files (using R)
- chapter 8: download and analyze SEC quarterly index files (using Perl)
- chapter 9: download SEC filings using R
- chapter 10: download SEC filings using Perl
- chapter 11: number of lines for 10-K (using R)
- chapter 12: number of lines for 10-K (using Perl)
- chapter 13: 10-K (using R)
- chapter 14: 10-K (using Perl)
- chapter 15: 13-f using R
- chapter 16: 13-f using Perl
- chapter 17: 10-Q using R
- chapter 18: 10-Q using Perl
- chapter 19: Fog index and other readability measures using R
- chapter 20: Fog index and other readability measures using Perl
- chapter 21: Zipf's law
- chapter 22: positive, negative words (R)
- chapter 23: positive, negative words (Perl)
- chapter 24: Replicate some seminal papers
- chapter 25: R and MySQL
- chapter 26: Perl and MySQL
- chapter 27: list of potential term projects



Text analysis on SEC filings
(CRN: xxx, Fall2018xxx)

Instructor:	Paul Yan
Contact Information:	Email: yany@canisius.edu Phone: (716) 888-2604 Office: CT308
Lecture:	xxx @ OM111 (Financial Services Lab)
Office hours:	Xxx @ CT308
Prerequisites:	a) Business Analytics using R (FIN456A, MBA674A) b) one finance/accounting course: FIN311 (Corporate Finance) and Investments (FIN312) , or equivalents.
Textbook:	Added later
Websites:	<i>An Introduction to R</i> http://canisius.edu/~yany/doc/R-intro.pdf <i>The R Language Definition</i> http://canisius.edu/~yany/doc/R-lang.pdf <i>My related R web site:</i> http://canisius.edu/~yany/R.shtml
QR codes for two more web sites	https://www.r-project.org/ http://www3.canisius.edu/~yany/R.shtml  Fun question: how many lines of R codes to generate one image?
One-line R codes	<pre>> source("http://canisius.edu/~yany/course.R")</pre> Note: I will explain this line in week 2
Course Description:	There are several reasons why we should offer such a course. First, the unstructured information has a lion share of all information, 70% to 80% and it is reported that 80% of structured information came from unstructured one. Second, SEC filings is an important source of information (gold mine) since public companies, certain insiders, and broker-dealers are required to make regular SEC filings. Third, from SEC filings we could retrieve both structured information, such as annual sales and net income, and unstructured information such as MD&A (Management Discussion and Comments). Fourth, SEC filings

	could be downloaded free of charge. Fifth, the tools used are free as well. Sixth, via replication of several seminal papers, students would know how hard it is to collect and process the unstructured information and appreciate the benefits generated from unstructured information.
Capacity of the class	20 (ideally the number of students should be less than 15 since hands-on is critical)
Three objectives:	<ol style="list-style-type: none"> 1) Learn/review basic financial concepts such as Ratio Analysis, Portfolio Theory, CAPM, Fama-French-Carhart Factor Model, Monte Carlo simulation, Options Theory, VaR (Value at Risk) and Market microstructure 2) Learn and apply R to finance 3) Focus on publicly available financial data such as Yahoo Finance, Google Finance, Prof. French's Data Library and Federal Reserve Economic Data Library (FRED).
Perl software	Perl program: http://www.activestate.com/ DzSoft Perl editor: http://dzsoft.com
R Software	R is open source statistical and computational software https://www.r-project.org/
Academic Integrity:	Students are expected to know and understand college policies with regard to Academic Integrity Code . Violations of academic integrity will be prosecuted fully. Please note that you are responsible for reporting any instances where other students have violated these policies. Failure to do so will result in penalties as well. If you have any questions about this policy, please see the instructor.
Attendance Policy:	Attending classes regularly is required. Before-class preparation and in-class participation is an integral part of this course. Students are strongly encouraged to participate in class discussions and ask questions. Students are encouraged to discuss current events relevant to this course or their own experiences. Homework problems are regularly assigned.
Academic and Accessibility Support Services:	The GRIFF Center for Academic Engagement provides comprehensive programs, tutoring services, and resources to support student academic and career success. If you would like to learn more about academic support, please stop in Old Main 013 or call 716-888-2170. Visit the GRIFF Center webpage at: http://www.canisius.edu/griff-center/ . Accessibility Support (716-888-2170), which is located in the Griff Center for Academic Engagement (OM 013), is responsible for arranging appropriate academic accommodations for students with documented disabilities. If anyone in this course falls into this category, please contact Accessibility Support so that an appropriate course of action may be determined. For additional information, see http://www.canisius.edu/dss/
Course Level Learning Goals:	Learn basic finance theory; understand various decision rules, such as NPV, IRR and payback rules to evaluate different investment projects; understand the term-structure of interest rate, spread, and default risk; learn how to download data from public sources, such as

	Yahoo!Finance, Federal Reserve Bank’s Data Library, Prof. French’s Data Library, be able to use Excel to process data and calculate certain measures, .e.g., estimate the market risk for a given set of tickers by applying CAPM.																
College, Program and Major Learning Goals:	This course is designed to help students achieve one or more College Core, Business Program and/or Major level learning goals and objectives. You can see the specific College, Program or Major level learning goals and objectives associated with the course from this page on the College website: http://bit.ly/bcoreLG																
Grade Evaluation:	<table> <tr> <td>Homework</td> <td>30%</td> </tr> <tr> <td>Midterm</td> <td>20%</td> </tr> <tr> <td>Final exam</td> <td>25%</td> </tr> <tr> <td>Group project</td> <td>10%</td> </tr> <tr> <td>Group presentation</td> <td>5%</td> </tr> <tr> <td>Class participation</td> <td>10%</td> </tr> <tr> <td colspan="2">-----</td> </tr> <tr> <td>Total</td> <td>100%</td> </tr> </table>	Homework	30%	Midterm	20%	Final exam	25%	Group project	10%	Group presentation	5%	Class participation	10%	-----		Total	100%
Homework	30%																
Midterm	20%																
Final exam	25%																
Group project	10%																
Group presentation	5%																
Class participation	10%																

Total	100%																
Teaching Methods:	Each class will be consist of two parts: lecture (including discussion of homework) and hands-on.																
Group project	<p>Each group can have up to three members. A topic should be closely associated with this course. The maximum number of pages of your report is 15 with 12-point font. Please discuss with me your topic before you start to work on it. Some basic criterions are listed below. Real world topics are especially encouraged. Three parts are essential:</p> <ol style="list-style-type: none"> 1) theory and background of the topic, 2) R programs with a short explanation of the codes, 3) final data set (plus the codes to process the data, the source of raw data) <p>Note: please do not send me your raw data.</p> <p>The second type of projects is to study one of R packages. Three parts are critical:</p> <ol style="list-style-type: none"> 1)why this specific package is useful in finance 2)a summary of all or most important functions offered by the package 3)examples to use them <p>Note: see a list of potential topics, at the end of the syllabus, for the group projects.</p>																
Makeup Exams:	To be eligible for a makeup examination, a student must submit to the instructor written documentation of the reason for missing a scheduled examination due to medical problems or death of an immediate family member. This has to be done before the scheduled exam. The instructor (not the student) determines whether and when a makeup is to be given. If a makeup examination is to be given, the instructor will determine the type of makeup examination. If the student misses (for any reason) the scheduled makeup examination, additional makeups are not permissible.																

Laptop policy:	Students are encouraged to bring a laptop to the class. However, it should be used for class related activities only.
Course Schedule:	For the detailed schedule, see below. I reserve the right to change the course schedule throughout the semester. Changes to the schedule will be announced in class or via email.

Tentative schedule

#	Date	Topics	Description (F for Finance)
1	8/30 9/1	Syllabus discussion, introduction to R and Perl	A short survey, self-intro, syllabus, course structure, mid-term and final R: Installation, assignment, basic math functions: mean(), min(),max(), median(), sd() and use R as a scientific calculator P: Perl installation, DzSoft Perl editor
2	9/6 9/8	R basics	S: Introduction to SEC filings R: R basics How to write an R function? double_f(), pv_f(), fv_f(), IRR(), How to call your functions? several ways to input data, matrix, differential operator and use R as a financial calculator
3	9/13 9/15	Perl basics	P: R basics Perl: Perl basics
4	9/20 9/22	SEC index	R: download one quarterly index, download all quarterly indices Perl: download one quarterly index file, all index files
5	9/27 9/29	Analyze index files	R: Perl
6	10/4 10/6	Download text filings (files)	R: Perl:
7	10/11 10/13	Regular expression	R: Perl:
8	10/18 11/20	10-K	Midterm
9	10/25 10/27	10-K	R: t.test(), var.test(), dwtest(), Wilcoxon.test(), granger_test(). Perl: Note: post about a dozen topics in detail
10	11/1 11/3	13-f	
11	11/8 11/10	N-STAR	R: stockMonthly, indexMonthly, indexDaily, stockD1925 to stockD2014, various R program to retrieve/process data efficiently
12	11/15 11/17	10-Q	
13	11/22 11/24	Other forms	Thanks giving holiday (no Class)
14	11/29 12/1	Group Presentation	3 to 4 groups
15	12/6 12/8	Group Presentation	3 to 4 groups All other groups
	TBA	Final	Final-exam

Appendix I: Instruction to replicate Li (2008)

Objectives of this exercise

- a) understand SEC filings
- b) know how to download 10-K filings by using Perl
- c) understand Perl's regular regression
- d) understand MD&A (Management Discussion and Analysis)
- e) understand the definition of Fog index
- f) try about several dozen MD&A and generate related Fog index

$$\text{Fog index} = 0.4 * (n + p) \quad (1)$$

where, n is the average number of words per sentence, while p is percentage of complex words. A complex word is a word has more than two syllables.

Abstract of Li (2008)

This paper examines the relation between annual report readability and firm performance and earnings persistence. I measure the readability of public company annual reports using the Fog index from the computational linguistics literature and the length of the document. I find that: (1) the annual reports of firms with lower earnings are harder to read (i.e., they have a higher Fog index and are longer); and (2) firms with annual reports that are easier to read have more persistent positive earnings.

From Li(2008)

Appendix A. Steps to calculate the readability indices

This appendix explains the details of calculating the readability indices starting from the raw 10-K filings used in this paper. I first download the 10-K report from Edgar and perform the following editing before further analysis. First, the heading information that is contained between <SEC-HEADER> and <SEC-HEADER> is deleted. Second, all the tables that begin with <TABLE> and end with <TABLE> or the paragraphs that contain <S> or <C> are deleted, because <S> and <C> tags are used by some firms to present tables. Next, all the tags in the format of <. . .> and <&...>, which are used widely in documents in SEC HTML or XML format documents, are replaced with blanks. Finally, to make sure that all the tables, tabulated text, or financial statements are excluded, all the paragraphs with more than 50% of non-alphabetic characters (e.g., white spaces or numbers) are deleted.

The file after the editing is then analyzed using the Fathom package in Perl. The package can calculate the typical text statistics, including the number of characters, number of words, percent of complex words (i.e., words with more than three syllables), number of sentences, number of text lines, number of paragraphs, syllables per word, and words per sentence. Based on the statistics, the package also produces the summary readability indices used in the paper.

From Li (2008), Appendix B. Steps to extract MD&A and Notes to the financial statements This appendix explains the details of extracting the MD&A section and Notes from 10-K filings. Starting with the raw 10-K file, I first delete the SEC-header information, all the contents between <TABLE> and <TABLE> text, the paragraphs that contain <S> or <C>, all

the tags in the format of < . . .> and <&...> are removed using the same process described in Appendix A.

Within the remaining text, the program identifies a line that satisfies one of the following criteria as the beginning of the MD&A section: (1) the line starts with “management’s discussion” or “management’s discussion” following some white spaces; (2) the line contains “management’s discussion” and (“item”+one or more white space+“7”) and does not contain the word “see”; (3) the line starts with some white spaces followed by “managements discussion” or “managements discussion”; or (4) the line contains “managements discussion” and (“item”+one or more white space+“7”) and does not contain the word “see.” Since many firms refer to the MD&A section in the front-matter of the annual reports, the word “see” serves to identify all such situations. The program identifies a line that satisfies one of the following criteria as the ending of the MD&A section: (1) the line begins with some white spaces followed by “Financial Statements” or “Financial Statements”; (2) the line contains “item” followed by one or more white spaces and the number “8”; (3) the line contains “Supplementary Data”; or (4) the line begins with some white spaces followed by “SUMMARY OF SELECTED FINANCIAL DATA” or “SUMMARY OF SELECTED FINANCIAL DATA.” Most firms have a table of contents listing the main sections of the 10-K filing. In some instances, this table of contents is not embedded between <TABLE> and <TABLE> and therefore is not cleaned in the previous steps. As a result, the line in the table of contents about MD&A will also be picked up by the program as part of the MD&A.

Similarly, the program identifies a line as the beginning of the Notes, if: (1) the line starts with “NOTES TO” or some white spaces followed by “NOTES TO”; and (2) the line does not contain any number except when it follows “for the years ended.” The program identifies a line that satisfies one of the following criteria as the ending of the Notes: (1) the line contains “Changes in and Disagreements with Accountants” or “DISAGREEMENTS ON ACCOUNTING”; (2) the line contains “DIRECTORS AND EXECUTIVE OFFICERS”; or (3) the line contains “exhibit index.”

After the MD&A and the Notes are identified, all the paragraphs with more than 50% of non-alphabetic characters (e.g., white spaces or numbers) are deleted. Finally, the Fathom package is used to calculate the readability measures.

Appendix J: a list of index files from SEC filings

 index19931.zip	5/14/2016 7:58 PM	Compressed (zipp...	1 KB
 index19932.zip	5/14/2016 7:58 PM	Compressed (zipp...	1 KB
 index19933.zip	5/14/2016 7:58 PM	Compressed (zipp...	1 KB
 index19934.zip	5/14/2016 7:58 PM	Compressed (zipp...	2 KB
 index19941.zip	5/14/2016 7:58 PM	Compressed (zipp...	296 KB
 index19942.zip	5/14/2016 7:58 PM	Compressed (zipp...	249 KB
 index19943.zip	5/14/2016 7:58 PM	Compressed (zipp...	213 KB
 index19944.zip	5/14/2016 7:58 PM	Compressed (zipp...	233 KB
 index19951.zip	5/14/2016 7:58 PM	Compressed (zipp...	445 KB
 index19952.zip	5/14/2016 7:58 PM	Compressed (zipp...	413 KB
 index19953.zip	5/14/2016 7:58 PM	Compressed (zipp...	437 KB
 index19954.zip	5/14/2016 7:58 PM	Compressed (zipp...	470 KB
 index19961.zip	5/14/2016 7:58 PM	Compressed (zipp...	716 KB
 index19962.zip	5/14/2016 7:58 PM	Compressed (zipp...	790 KB
 index19963.zip	5/14/2016 7:58 PM	Compressed (zipp...	852 KB
 index19964.zip	5/14/2016 7:58 PM	Compressed (zipp...	899 KB
 index19971.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,323 KB
 index19972.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,061 KB
 index19973.zip	5/14/2016 7:58 PM	Compressed (zipp...	993 KB
 index19974.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,038 KB
 index19981.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,537 KB
 index19982.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,191 KB
 index19983.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,091 KB
 index19984.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,064 KB
 index19991.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,535 KB
 index19992.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,272 KB
 index19993.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,161 KB
 index19994.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,158 KB
 index20001.zip	5/14/2016 7:58 PM	Compressed (zipp...	1,718 KB
 index20002.zip	5/14/2016 7:59 PM	Compressed (zipp...	1,321 KB

Appendix K: A few files for 2014 Q3

Name	Date modified	Type	Size
 1000032_0001181431-14-030779.txt	6/8/2016 6:20 PM	Text Document	5 KB
 1000045_0001000045-14-000005.txt	6/10/2016 10:56 PM	Text Document	6 KB
 1000045_0001000045-14-000006.txt	6/10/2016 10:56 PM	Text Document	7 KB
 1000045_0001085146-14-001812.txt	6/10/2016 10:56 PM	Text Document	31 KB
 1000045_0001193125-14-257879.txt	6/10/2016 10:56 PM	Text Document	162 KB
 1000045_0001193125-14-264054.txt	6/10/2016 10:56 PM	Text Document	622 KB
 1000045_0001193125-14-297095.txt	6/10/2016 10:56 PM	Text Document	100 KB
 1000045_0001193125-14-304053.txt	6/10/2016 10:56 PM	Text Document	3,142 KB
 1000045_0001193125-14-312657.txt	6/10/2016 10:56 PM	Text Document	24 KB
 1000097_0000919574-14-004836.txt	6/10/2016 10:03 AM	Text Document	95 KB
 1000177_0000000000-14-040049.txt	6/10/2016 11:05 PM	Text Document	195 KB
 1000177_0000000000-14-042163.txt	6/10/2016 11:05 PM	Text Document	77 KB
 1000177_0000919574-14-004886.txt	6/10/2016 11:05 PM	Text Document	28 KB
 1000177_0001193125-14-285718.txt	6/10/2016 11:05 PM	Text Document	749 KB
 1000180_0001000180-14-000043.txt	6/11/2016 3:45 PM	Text Document	536 KB
 1000180_0001000180-14-000050.txt	6/11/2016 3:44 PM	Text Document	18,242 KB
 1000180_0001000180-14-000054.txt	6/11/2016 3:45 PM	Text Document	44 KB
 1000180_0001000180-14-000058.txt	6/11/2016 3:45 PM	Text Document	44 KB
 1000180_0001104659-14-051240.txt	6/11/2016 3:45 PM	Text Document	133 KB
 1000180_0001104659-14-051741.txt	6/11/2016 3:45 PM	Text Document	135 KB
 1000180_0001104659-14-052046.txt	6/11/2016 3:45 PM	Text Document	116 KB
 1000180_0001104659-14-052855.txt	6/11/2016 3:45 PM	Text Document	123 KB
 1000180_0001104659-14-052856.txt	6/11/2016 3:45 PM	Text Document	44 KB
 1000180_0001104659-14-052983.txt	6/11/2016 3:45 PM	Text Document	2,473 KB
 1000180_0001242648-14-000059.txt	6/11/2016 3:44 PM	Text Document	6 KB