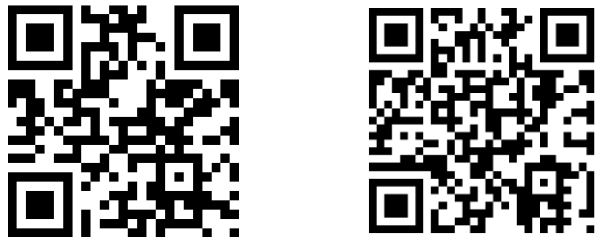




Text analysis on SEC filings
(CRN: xxx, Fall2018xxx)

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 equivalent.
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	> <i>source("http://canisius.edu/~yany/course.R")</i> 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: numbers of records, number of unique forms, frequency of each form Perl: numbers of records, number of unique forms, frequency of each form
6	10/4 10/6	Download text filings (files)	R: how to generate appropriate forms Perl: [same]
7	10/11 10/13	Word frequency tables	R and Perl: clean data by removing <Table> <S> etc, word frequencies and word sequence frequencies wordfrequency.rda, LMlist.rda (Lougrahm and McDonald 2011), posNegWords.rds
8	10/18 11/20	Regular expression and 10-K parsing	R and Perl: how to parse 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	Perl: download the process 13-f Shareholdings for financial institutions
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	Parse quarterly financial statements
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

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.
- Leone, Andy, 2014 Web-Based Data Collection and Analysis
- 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, 2015, Textual Analysis in Accounting and Finance: A survey, working paper, University of Notre Dame.
- Loughran, Tim, Bill McDonald, 2014, Measuring Readability in Financial Disclosures, *Journal of Finance*,
- 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.