1 Related Work

The collection of papers with detailed notes can be found here:
https://drive.google.com/drive/folders/1GSDjs8FwtoIpJVbJ-g2lYR7QRO-MO4xB?usp=sharing

Ref.	Title	Score	Comments
Surveys			
[51]	*(2021) A Survey on Extraction of Causal Relations from Natural Language Text*	**10/10**	causality extraction, good comparison and summary, structured view; useful information on dataset (pros/cons + limitations)
[43]	*(2021) Text analysis in financial disclosures*	**10/10**	good background info; very relevant; reference comparison embedded in text (not in tables)
[11]	*(2021) Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond*	**4/10**	causal inference, not very relevant; deducing cause and effect for social science, not concrete identification and extraction of causal phrases.
[14]	*(2020) Comprehensive review of text-mining applications in finance*	**3/10**	badly written, not very comprehensive, some interesting references
[48]	*(2019) Text Mining in Financial Information*	**9/10**	poorly written, summary table very useful references
[38]	*(2019) Text Mining for Big Data Analysis in Financial Sector: A Literature Review*	**7/10**	some useful pointers, interesting methodology of identifying key research paper - network analysis
[36]	*(2019) Identification of Causal Dependencies by Using Natural Language Processing: A Survey*	**5/10**	not very comprehensive/concrete; some reference might be useful
[13]	*(2016) Textual analysis and machine leaning: Crack unstructured data in finance and accounting*	**4/10**	very basic, a bit outdated; mainly classification and LDA analysis
Support			
[26]	*(2019) Fad or future? Automated analysis of financial text and its implications for corporate reporting*	**10/10**	very RELEVANT! addresses key issues with critical reflection; management/accounting perspective; good writing style
[53]	*(2018) Causal language intensity in performance commentary and financial analyst behaviour*	**10/10**	very RELEVANT; provide a support base for further studies
[24]	*(2018) About relationship between business text patterns and financial performance in corporate data*	**9/10**	helpful conclusions re word counts and sentiment; useful visualisation for EDA
[25]	*(2017) Identifying Emerging Trends of Financial Business Method Patents*	**4/10**	interesting approach of identifying emerging trends, questionable relevance - mainly topic modelling / LDA; survial analysis, hot vs. cold
[8]	*(2016) Learning tone and attribution for financial text mining*	**7/10**	useful list of key words; basic; some good reference; public dataset for benchmarking?
[7]	*(2012) Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language*	**7/10**	important word list, prior research theories
[16]	*(2007) Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm*	**9/10**	a bit outdated but provides theoretical support; useful thesaurus list
[23]	*(2002) Combining data and text mining techniques for analysing financial reports*	**8/10**	outdated but interesting SOM algorithm, sort of in the right direction

Extraction / Semantic Parsing			
[49]	*(2021) TDJEE: A Document-Level Joint Model for Financial Event Extraction*	**8/10**	event extraction, template-slot filling, Chinese financial docs, restricted to 5 types of PRE-DEFINED events
[6]	*(2020) Architecture and Evolution of Semantic Networks in Mathematics Texts*	**9/10**	linear algebra text book, pre-defined terms, concept gaps, core vs peripheral
[33]	*(2021) Mind Map Automation: Using Natural Language Processing to Graphically Represent a Portion of a U.S. History Textbook*	**5/10**	interesting application history text book, but very simplistic - BERT summarization in original sentence
[39]	*(2019) A Study on Semantic Parsing of Cooking Recipes*	**9/10**	good elaboration on background concepts and models ('dependency parser', 'NER', 'Frame and slots', etc.); strong introduction but weaker body of content; annotation schema not relevant
[9]	*(2020) Financial Knowledge Graph Construction*	**7/10**	
[10]	*(2020) A High Precision Pipeline for Financial Knowledge Graph Construction*	**7/10**	PhD thesis and a shorter version in a publish paper; interesting pipeline and implementation details; useful lexicon terms and bigrams; events mostly on announcement, appointment, M&A, etc.
[19]	*(2021) Exploring Construction of a Company Domain-Specific Knowledge Graph from Financial Texts Using Hybrid Information Extraction*	**[8]/10**	Master's thesis. Company-specific entity relationships; hybrid approach (rule-based lexical patterns + pre-trained); to be looked into details
[44]	*(2017) Developing a Methodology of Structuring and Layering Technological Information in Patent Documents through Natural Language Processing*	**9/10**	useful framework, pointing word, similarity (patent vs MDA)
[17]	*(2014) Automated Detection of Financial Events in News Text*	**[.]/10**	PhD thesis. Very detailed.
[46]	*(2013) Automatic extraction of cause-effect relations in natural language text*	**9/10**	useful logic rules and linguistic patterns for recognising cause and effect; could be adopted in combination with weak supervision
[55]	*(2012) The automatic creation of concept maps from documents written using morphologically rich languages*	**7/10**	relevant idea but quite generic; no implementation details; no visual output; non-English
[5]	*(2005) Processing Complex Sentences for Information Extraction*	**7/10**	outdated but relevant; similar approach / workflow; application area - biomedical text; poor quality of writing
[37]	*(2004) Newsmap: A knowledge map for online news*	**8/10**	SOM (self-organizing map) algorithm; phrase extraction via frequency reduction; layered map is also interesting
[1]	*(1992) Automatic Extraction of Facts from Press Releases to Generate News Stories*	**7/10**	Old but relevant; patterns to fill slots; useful evaluation metric
Text Classification			

[27]	*(2020) A Unified Model for Financial Event Classification, Detection and Summarization*	**6/10**	basic jointly learning model; end application not very clear; small dataset (1600 docs) and manual annotation
[47]	*(2020) LEDGAR: A Large-Scale Multi-label Corpus for Text Classification of Legal Provisions in Contracts*	**8/10**	text classification based on EDGAR documents, legal text, financial documents, code base available
[15]	*(2015) Financial Footnote Analysis: Developing a Text Mining Approach*	**6/10**	not very relevant; pre-defined categories on income tax notes, limited scope, very basic classification
Knowledge Graph			
[50]	*(2021) Data Set and Evaluation of Automated Construction of Financial Knowledge Graph*	**[.]/10**	to come
[30]	*(2019) Anticipating Stock Market of the Renowned Companies: A Knowledge Graph Approach*	**8/10**	applying KG to extract features - TransE a bit 'forced'; results not very impressive; some useful discussion of limitations
[31]	*(2018) Stock Price Movement Prediction from Financial News with Deep Learning and Knowledge Graph Embedding*	**8/10**	Similar idea as above; same first author, published one year earlier
[4]	*(2018) Mapping Deep NLP to Knowledge Graphs: An Enhanced Approach to Analyzing Corporate Filings with Regulators*	**9/10**	SEC filings to KG construction; focused on directors/executives; no visualization of output
[28]	*(2018) Combining Enterprise Knowledge Graph and News Sentiment Analysis for Stock Price Volatility Prediction*	**7/10**	KG focused on shareholding, cooperation, management, and supply-customer. sentiment dictionary, Chinese news
Background			
[45]	*(2021) Leveraging Text Mining and Analytical Technology to Enhance Financial Planning and Analysis*	**3/10**	some useful background information, very 'watery' MBA non-finance, not computer science, no programming, not impressive at all.
[2]	*(2018) OpenEDGAR: Open Source Software for SEC EDGAR Analysis*	**9/10**	GOOD background information on EDGAR
[54]	*(2020) Progress in Neural NLP: Modeling, Learning, and Reasoning*	**9/10**	comprehensive basic NLP textbook material
[32]	*(2018) Beyond concept analysis: Uses of mind mapping software for visual representation, management, and analysis of diverse digital data*	**1/10**	Not useful; content not relevant; interesting framework: add, analyze, revise; definition of mindmap - spatial
[35]	*(2018) "CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures"*	**1/10**	not relevant; semantic annotation schema; toy dataset - 320 short stories
[22]	*(2010) Investigating Causal Relationships in Stock Returns with Temporal Logic Based Methods*	**4/10**	not based on text analytics just time series of price data; interesting propositional probabilistic temporal logic form; dubious assumptions re normal distribution. not very relevant
[3]	*(2010) Link Analysis in Mind Maps: A New Approach To Determine Document Relatedness*	**1/10**	not relevant

[21]	*(2019) Reasoning-Driven Question-Answering For Natural Language Understanding*	**[.]/10**	comprehensive coverage of reasoning, formalism and NLU
Inspiration			
[41]	*(2016) Data Programming: Creating Large Training Sets, Quickly*	**10/10**	original paper
[40]	*(2017) Snorkel: rapid training data creation with weak supervision*	**10/10**	subsequent system implementation
[42]	*(2019) Accelerating machine learning with training data management*	**10/10**	PhD Thesis with detailed explanation
[52]	*(2020) Weakly Supervised Subevent Knowledge Acquisition*	**10/10**	interesting framework
Deep Reinforcement Learning for Trading			
[12]	*(2018) Reinforcement learning in financial markets - a survey*	**10/10**	survey
[20]	*(2021) Applications of deep learning in stock market prediction: Recent progress*	**10/10**	survey, very comprehensive and relevant; good coverage and comparison overview
[29]	*(2020) FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance*	**10/10**	useful resources / libraries for implementing RL trading workflow
[18]	*(2020) Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy*	**9/10**	good references, very relevant, only price information
[34]	*(2020) Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics*	**2/10**	badly written, seems like study notes rather than published paper, some good collection of fundamental concepts

References

[1] Peggy M. Andersen, Philip J. Hayes, Steven P. Weinstein, Alison K. Huettner, Linda M. Schmandt, and Irene B. Nirenburg. Automatic extraction of facts from press releases to generate news stories. In *Third Conference on Applied Natural Language Processing*, pages 170–177, Trento, Italy, March 1992. Association for Computational Linguistics.

[2] Michael J Bommarito II au2, Daniel Martin Katz, and Eric M Detterman. Openedgar: Open source software for sec edgar analysis, 2018.

[3] Joeran Beel and Bela Gipp. Link analysis in mind maps: A new approach to determine document relatedness. 01 2010.

[4] Damir Cavar and Matthew Josefy. Mapping deep nlp to knowledge graphs: An enhanced approach to analyzing corporate filings with regulators. In Mahmoud El-Haj, Paul Rayson, and Andrew Moore, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, may 2018. European Language Resources Association (ELRA).

[5] Deepthi Chidambaram. Processing complex sentences for information extraction. Master's thesis, Arizona State University, 2015.

[6] Nicolas H. Christianson, Ann Sizemore Blevins, and Danielle S. Bassett. Architecture and evolution of semantic networks in mathematics texts. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2239):20190741, Jul 2020.

[7] Angela Davis, Jeremy Piger, and Lisa Sedor. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 29:845–868, 09 2012.

[8] Mahmoud El-Hay, Paul Rayson, Steve Young, Martin Walker, Thomas Schleicher, and Vasiliki Athanasakou. Learning tone and attribution for financial text mining. In *In: N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.) Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA)*, page 1820, 2016.

[9] Sarah Elhammadi. *Financial knowledge graph construction.* PhD thesis, University of British Columbia, 2020.

[10] Sarah Elhammadi, Laks V.S. Lakshmanan, Raymond Ng, Michael Simpson, Baoxing Huai, Zhefeng Wang, and Lanjun Wang. A high precision pipeline for financial knowledge graph construction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 967–977, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[11] Amir Feder, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond, 2021.

[12] Thomas G. Fischer. Reinforcement learning in financial markets - a survey. In *FAU Discussion Papers in Economics 12/2018, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics*, 2018.

[13] Feng i Guo. Textual analysis and machine leaning: Crack unstructured data in finance and accounting. *The journal of finance and data science.*, 2(3).

[14] A. Gupta, V. Dengre, and H.A. Kheruwala. Comprehensive review of text-mining applications in finance, 2020.

[15] Maryam Heidari and Carsten Felden. Financial footnote analysis: Developing a text mining approach. In *International Conference Data Mining (DMIN 15)*, 2015.

[16] Elaine Henry. Market reaction to verbal components of earnings press releases: Event study using a predictive algorithm. *Journal of Emerging Technologies in Accounting*, 3, 01 2007.

[17] Frederik Hogenboom. *Automated Detection of Financial Events in News Text*. PhD thesis, Erasmus Research Institute of Management (ERIM), 12 2014.

[18] Shan Zhong Hongyang Yang, Xiao-Yang Liu and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *ICAIF '20: ACM International Conference on AI in Finance, Oct. 15–16, 2020, Manhattan, NY. ACM, New York, NY, USA.*, 2020.

[19] Chun-Heng Jen. Exploring construction of a company domain-specific knowledge graph from financial texts using hybrid information extraction. Master's thesis, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, 2021.

[20] Weiwei Jiang. Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 184:115537, Dec 2021.

[21] Daniel Khashabi. *Reasoning-Driven Question-Answering For Natural Language Understanding*. PhD thesis, University of Pennsylvania, 2019.

[22] Samantha Kleinberg, Petter Kolm, and Bud Mishra. Investigating causal relationships in stock returns with temporal logic based methods. 06 2010.

[23] Antonina Kloptchenko, Tomas Eklund, Jonas Karlsson, Barbro Back, Hannu Vanharanta, and Ari Visa. Combining data and text mining techniques for analysing financial reports. *International Journal of Intelligent Systems in Accounting*, 12, 01 2002.

[24] BangRae Lee, Jun-Hwan Park, Leenam Kwon, Young-Ho Moon, Youngho Shin, GyuSeok Kim, and Hanjoon Kim. About relationship between business text patterns and financial performance in corporate data. *Journal of Open Innovation: Technology, Market, and Complexity*, 4, 12 2018.

[25] Won Sang Lee and So Young Sohn. Identifying emerging trends of financial business method patents. *Sustainability*, 9(9), 2017.

[26] Craig M. Lewis and Steven Young. Fad or future? automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, 49:587 – 615, 2019.

[27] Quanzhi Li and Qiong Zhang. A unified model for financial event classification, detection and summarization. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4668–4674. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Special Track on AI in FinTech.

[28] Jue Liu. Combining enterprise knowledge graph and news sentiment analysis for stock price volatility prediction. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2018.

[29] Xiao-Yang Liu, Hongyang Yang, Qian Chen, Runjia Zhang, Liuqing Yang, Bowen Xiao, and Christina Dan Wang. Finrl: A deep reinforcement learning library for automated stock trading in quantitative finance, 2020.

[30] Yang Liu, Qingguo Zeng, Joaquín B. Ordieres Meré, and Huanrui Yang. Anticipating stock market of the renowned companies: A knowledge graph approach. *Hindawi Complexity*, 2019:9202457:1–9202457:15, 2019.

[31] Yang Liu, Qingguo Zeng, Huanrui Yang, and Adrian Carrio. *Stock Price Movement Prediction from Financial News with Deep Learning and Knowledge Graph Embedding*, pages 102–113. 07 2018.

[32] Jennifer Mammen and Corey Mammen. Beyond concept analysis: Uses of mind mapping software for visual representation, management, and analysis of diverse digital data. *Research in Nursing & Health*, 41, 11 2018.

[33] Sophie McIntyre. Mind map automation: Using natural language processing to graphically represent a portion of a u.s. history textbook. Master's thesis, University Honors College Middle Tennessee State University, 2021.

[34] Amir Mosavi, Pedram Ghamisi, Yaser Faghan, Puhong Duan, and Shahab Shamshirband. Comprehensive review of deep reinforcement learning methods and applications in economics. Mar 2020.

[35] Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende. CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. In *Proceedings of the Fourth Workshop on Events*, pages 51–61, San Diego, California, June 2016. Association for Computational Linguistics.

[36] Erika Nazaruka. Identification of causal dependencies by using natural language processing: A survey. In *Proceedings of the 14th International Conference on Evaluation of Novel Approaches to Software Engineering*, ENASE 2019, page 603–613, Setubal, PRT, 2019. SCITEPRESS - Science and Technology Publications, Lda.

[37] Thian-Huat Ong, Hsiu-chin Chen, Wai-Ki Sung, and Bin Zhu. Newsmap: A knowledge map for online news. *Decision Support Systems*, 39:583–597, 05 2004.

[38] Mirjana Pejić Bach, Živko Krstić, Sanja Seljan, and Lejlaf Turulja. Text mining for big data analysis in financial sector: A literature review. *Sustainability*, 11(5), 2019.

[39] Florian Pfisterer. A study on semantic parsing of cooking recipes, November 2019.

[40] Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: rapid training data creation with weak supervision. *Proceedings of the VLDB Endowment*, 11(3):269–282, Nov 2017.

[41] Alexander Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. Data programming: Creating large training sets, quickly. *Advances in neural information processing systems*, 29, 05 2016.

[42] Alexander Jason Ratner. *Accelerating machine learning with training data management*. PhD thesis, Stanford University, 2019.

[43] Sridhar Ravula. Text analysis in financial disclosures, 2021.

[44] Taeyeoun Roh, Yujin Jeong, and Byungun Yoon. Developing a methodology of structuring and layering technological information in patent documents through natural language processing. *Sustainability*, 9(11), 2017.

[45] Yih-Shan Sheu. Leveraging text mining and analytical technology to enhance financial planning and analysis. Master's thesis, Iowa State University, 2021.

[46] Antonio Sorgente, G. Vettigli, and Francesco Mele. Automatic extraction of cause-effect relations in natural language text. 1109:37–48, 01 2013.

[47] Don Tuggener, Pius von Däniken, Thomas Peetz, and Mark Cieliebak. LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1235–1241, Marseille, France, May 2020. European Language Resources Association.

[48] Nida Turegun. Text mining in financial information. 1:18–26, 01 2019.

[49] Peng Wang, Zhenkai Deng, and Ruilong Cui. Tdjee: A document-level joint model for financial event extraction. *Electronics*, 10(7), 2021.

[50] Wenguang Wang, Yonglin Xu, Chunhui Du, Yunwen Chen, Yijie Wang, and Hui Wen. Data Set and Evaluation of Automated Construction of Financial Knowledge Graph. *Data Intelligence*, 3(3):418–443, 09 2021.

[51] Jie Yang, Soyeon Caren Han, and Josiah Poon. A survey on extraction of causal relations from natural language text, 2021.

[52] Wenlin Yao, Zeyu Dai, Maitreyi Ramaswamy, Bonan Min, and Ruihong Huang. Weakly Supervised Subevent Knowledge Acquisition. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5345–5356, Online, November 2020. Association for Computational Linguistics.

[53] Shuyu Zhang, Walter Aerts, and Huifeng Pan. Causal language intensity in performance commentary and financial analyst behaviour. *Journal of Business Finance & Accounting*, 46, 08 2018.

[54] Ming Zhou, Nan Duan, Shujie Liu, and Heung-Yeung Shum. Progress in neural nlp: Modeling, learning, and reasoning. *Engineering*, 6(3):275–290, 2020.

[55] Krunoslav Zubrinic, Damir Kalpic, and Mario Milicevic. The automatic creation of concept maps from documents written using morphologically rich languages. *Expert Systems with Applications*, 39(16):12709–12718, Nov 2012.