

FNP 2019

**Proceedings of the Second Financial Narrative Processing  
Workshop (FNP 2019)**

30 September, 2019

Turku University

Turku, Finland

©2019 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-950737-19-2

## Preface

Welcome to the Second Financial Narrative Processing Workshop (FNP 2019) held at NoDaLiDa 2019 in Turku, Finland. Following the success of the First FNP 2018 at LREC'18, Japan, we have had a great deal of positive feedback and interest in continuing the development of the financial narrative processing field. This prompted us to hold a training workshop in textual analysis methods for financial narratives that was oversubscribed showing that there is an increasing interest in the subject. As a result, we were motivated to organise the Second Financial Narrative Processing Workshop, FNP 2019. The workshop will focus on the use of Natural Language Processing (NLP), Machine Learning (ML), and Corpus Linguistics (CL) methods related to all aspects of financial text mining and financial narrative processing (FNP). There is a growing interest in the application of automatic and computer-aided approaches for extracting, summarising, and analysing both qualitative and quantitative financial data. In recent years, previous manual small-scale research in the Accounting and Finance literature has been scaled up with the aid of NLP and ML methods, for example to examine approaches to retrieving structured content from financial reports, and to study the causes and consequences of corporate disclosure and financial reporting outcomes. One focal point of the proposed workshop is to develop a better understanding of the determinants of financial disclosure quality and the factors that influence the quality of information disclosed to investors beyond the quantitative data reported in the financial statements. The workshop will also encourage efforts to build resources and tools to help advance the work on financial narrative processing (including content retrieval and classification) due to the dearth of publicly available datasets and the high cost and limited access of content providers. The workshop aims to advance research on the lexical properties and narrative aspects of corporate disclosures, including glossy (PDF) annual reports, US 10-K and 10-Q financial documents, corporate press releases (including earning announcements), conference calls, media articles, social media, etc.

For FNP 2019 we collaborated with Fortia Financial Solutions, a French based company specialised in Financial Investment and Risk management who will work with us on organising a shared task on automatic detection of financial documents structure as part of FNP 2019. <http://fortia.fr/>

We accepted 11 submissions of which are 5 main workshop papers and 6 shared task papers, all papers accepted for oral presentation in the workshop. The papers cover a diverse set of topics in financial narratives processing reporting work on financial reports from different stock markets around the globe presenting analysis of financial reports. The quantity and quality of the contributions to the workshop are strong indicators that there is a continued need for this kind of dedicated Financial Narrative Processing workshop. We would like to acknowledge all the hard work of the submitting authors and thank the reviewers for the valuable feedback they provided. We hope these proceedings will serve as a valuable reference for researchers and practitioners in the field of financial narrative processing and NLP in general.

## **Organising Committee**

- Mahmoud El-Haj, Lancaster University, UK
- Paul Rayson, Lancaster University, UK
- Steven Young, Lancaster University, UK
- Houda Bouamor, Carnegie Mellon University in Qatar, Qatar
- Sira Ferradans, Fortia Financial Solutions, France
- Catherine Salzedo, Lancaster University, UK

## Program Committee

- Antonio Moreno Sandoval (UAM, Spain)
- Catherine Salzedo (LUMS, Lancaster University, UK)
- Denys Proux (Naver Labs, Switzerland)
- Djamé Seddah (INRIA-Paris, France)
- Eshrag Refaee (Jazan University, Saudi Arabia)
- George Giannakopoulos (SKEL Lab – NCSR Demokritos, Greece)
- Haithem Afli (Cork Institute of Technology, Ireland)
- Houda Bouamor (Carnegie Mellon University in Qatar, Qatar)
- Mahmoud El-Haj (SCC, Lancaster University, UK)
- Marina Litvak (Sami Shamoon College of Engineering, Israel)
- Martin Walker (University of Manchester, UK)
- Paul Rayson (SCC, Lancaster University, UK)
- Simonetta Montemagni (Istituto di Linguistica Computazionale – ILC, Italy)
- Sira Ferradans (Fortia Financial Solutions, France)
- Steven Young (LUMS, Lancaster University, UK)

## Preface

Welcome to the Second Financial Narrative Processing Workshop (FNP 2019) held at NoDaLiDa 2019 in Turku, Finland. Following the success of the First FNP 2018 at LREC'18, Japan, we have had a great deal of positive feedback and interest in continuing the development of the financial narrative processing field. This prompted us to hold a training workshop in textual analysis methods for financial narratives that was oversubscribed showing that there is an increasing interest in the subject. As a result, we were motivated to organise the Second Financial Narrative Processing Workshop, FNP 2019. The workshop will focus on the use of Natural Language Processing (NLP), Machine Learning (ML), and Corpus Linguistics (CL) methods related to all aspects of financial text mining and financial narrative processing (FNP). There is a growing interest in the application of automatic and computer-aided approaches for extracting, summarising, and analysing both qualitative and quantitative financial data. In recent years, previous manual small-scale research in the Accounting and Finance literature has been scaled up with the aid of NLP and ML methods, for example to examine approaches to retrieving structured content from financial reports, and to study the causes and consequences of corporate disclosure and financial reporting outcomes. One focal point of the proposed workshop is to develop a better understanding of the determinants of financial disclosure quality and the factors that influence the quality of information disclosed to investors beyond the quantitative data reported in the financial statements. The workshop will also encourage efforts to build resources and tools to help advance the work on financial narrative processing (including content retrieval and classification) due to the dearth of publicly available datasets and the high cost and limited access of content providers. The workshop aims to advance research on the lexical properties and narrative aspects of corporate disclosures, including glossy (PDF) annual reports, US 10-K and 10-Q financial documents, corporate press releases (including earning announcements), conference calls, media articles, social media, etc.

For FNP 2019 we collaborated with Fortia Financial Solutions, a French based company specialised in Financial Investment and Risk management who will work with us on organising a shared task on automatic detection of financial documents structure as part of FNP 2019. <http://fortia.fr/>

We accepted 11 submissions of which are 5 main workshop papers and 6 shared task papers, all papers accepted for oral presentation in the workshop. The papers cover a diverse set of topics in financial narratives processing reporting work on financial reports from different stock markets around the globe presenting analysis of financial reports. The quantity and quality of the contributions to the workshop are strong indicators that there is a continued need for this kind of dedicated Financial Narrative Processing workshop. We would like to acknowledge all the hard work of the submitting authors and thank the reviewers for the valuable feedback they provided. We hope these proceedings will serve as a valuable reference for researchers and practitioners in the field of financial narrative processing and NLP in general.

# Programme

09:00	09:15	Opening Remarks and Introduction to FNP
<b>Session 1</b>		<b>Financial Narrative Processing Papers</b>
09:15	09:40	Introduction to Financial Narrative Processing Tools and Resources Mahmoud El-Haj
09:40	10:05	Tone Analysis in Spanish Financial Reporting Narratives Antonio Moreno-Sandoval, Ana Gisbert, Pablo Alfonso Haya, Marta Guerrero and Helena Montoro
10:05	10:30	Coffee Break
10:30	10:55	Automated Stock Price Prediction Using Machine Learning Mariam Mokalled, Wassim El-Hajj and Mohamad Jaber
10:55	11:20	Utilizing Pre-Trained Word Embeddings to Learn Classification Lexicons with Little Supervision Frederick Blumenthal and Ferdinand Graf
11:20	11:45	Towards Unlocking the Narrative of the United States Income Tax Forms with Natural Language Processing Esme Manandise
11:45	12:10	Active learning for financial investment reports Sian Gooding and Ted Briscoe
12:10	13:30	Lunch Break
<b>Session 2</b>		<b>FinTOC Shared Task</b>
13:30	13:45	The FinTOC-2019 Shared Task: Financial Document Structure Extraction Remi Juge, Imane Bentabet and Sira Ferradans
13:45	14:00	UWB@FinTOC-2019 Shared Task: Financial Document Title Detection Tomas Hercig and Pavel Král
14:00	14:15	FinTOC-2019 Shared Task: Finding Title in Text Blocks Hanna Abi Akl, Anubhav Gupta and Dominique Mariko
14:15	14:30	FinDSE@FinTOC-2019 Shared Task Carla Abreu, Henrique Cardoso and Eugénio Oliveira

14:30	14:45	Daniel@FinTOC-2019 Shared Task : TOC Extraction and Title Detection Emmanuel Giguët and Gaël Lejeune
14:45	15:00	Finance document Extraction Using Data Augmented and Attention Ke Tian and Zi Jun Peng
15:00	15:05	Closing Remarks