

# **Second Workshop on Financial Technology and Natural Language Processing (FinNLP 2020)**

Kyoto, Japan  
5 January 2021

ISBN: 978-1-7138-2827-3

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

Financial News Annotation by Weakly-Supervised Hierarchical Multi-label Learning Hang Jiang, Zhongchen Miao, Yuefeng Lin, Chenyu Wang, Mengjun Ni, Jian Gao, Jidong Lu and Guangwei Shi.....	1
Variations in Word Usage for the Financial Domain Syrielle Montariol, Alexandre Allauzen and Asanobu Kitamoto.....	8
Detecting Omissions of Risk Factors in Company Annual Reports Corentin Masson and Syrielle Montariol.....	15
A Semantic Approach to Financial Fundamentals Jiafeng Chen and Suproteem Sarkar.....	22
Learning Company Embeddings from Annual Reports for Fine-grained Industry Characterization Tomoki Ito, Jose Camacho Collados, Hiroki Sakaji and Steven Schockaert.....	27
Unsupervised Discovery of Firm-Level Variables in Earnings Call Transcript Embeddings Daniel Edmiston and Ziho Park.....	34
Using Extractive Lexicon-based Sentiment Analysis to Enhance Understanding of the Impact of Non-GAAP Measures in Financial Reporting Stacey Taylor and Vlado Keselj.....	40
FinSBD-2020: The 2nd Shared Task on Sentence Boundary Detection in Unstructured Text in the Financial Domain Willy Au, Bianca Chong, Abderrahim Ait Azzi and Dialekti Valsamou-Stanislawski.....	47
PublishInCovid19 at the FinSBD-2 Task: Sentence and List Extraction in Noisy PDF Text Using a Hybrid Deep Learning and Rule-Based Approach Janvijay Singh.....	55
aiai at the FinSBD-2 Task: Sentence, list and Item Boundary Detection and Items classification of Financial Texts Using Data Augmentation and Attention Ke Tian, Hua Chen and Jie Yang.....	62
Daniel at the FinSBD-2 Task: Extracting List and Sentence Boundaries from PDF Documents, a model-driven approach to PDF document analysis Emmanuel Giguët and Gaël Lejeune.....	67
Subtl.ai at the FinSBD-2 task: Document Structure Identification by Paying Attention Abhishek Arora, Aman Khullar, Sarath Chandra Pakala, Vishnu Ramesh and Manish Shrivastava.....	75
The FinSim 2020 Shared Task: Learning Semantic Representations for the Financial Domain Ismail El Maarouf, Youness Mansar, Virginie Mouilleron and Dialekti Valsamou-Stanislawski.....	81

IITK at the FinSim Task: Hypernym Detection in Financial Domain via Context-Free and Contextualized Word Embeddings	
Vishal Keswani, Sakshi Singh and Ashutosh Modi.....	87
Anuj at the FinSim Task: Anuj@FINSIM–Learning Semantic Representation of Financial Domain with Investopedia	
Anuj Saini.....	93
ProsperAMnet at the FinSim Task: Detecting hypernyms of financial concepts via measuring the information stored in sparse word representations	
Gábor Berend, Norbert Kis-Szabó and Zsolt Szántó.....	98
FINSIM20 at the FinSim Task: Making Sense of Text in Financial Domain	
Vivek Anand, Yash Agrawal, Aarti Pol and Vasudeva Varma.....	104