

FinNLP 2019

**The First Workshop on Financial Technology and  
Natural Language Processing  
in conjunction with IJCAI 2019**

**Proceedings of the Workshop**

August 12, 2019

Macao, China

# Sponsor



科技部臺灣大學人工智慧研究中心

科技部人工智慧技術暨全幅健康照護聯合研究中心  
MOST JOINT RESEARCH CENTER FOR AI TECHNOLOGY AND  
ALL VISTA HEALTHCARE

## Preface

The aim of the Workshop on Financial Technology and Natural Language Processing (FinNLP workshop) is to provide a forum for international participants to share knowledge on applying NLP to the FinTech domain. With the sharing of the researchers in the FinNLP workshop, the challenging problems of blending FinTech and NLP will be identified, the future research directions will be shaped, and the scope of this interdisciplinary research area will be broadened.

The 1st FinNLP workshop is held in Macao on August 12, 2019, in conjunction with the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019). The participants came from both academia and industry. Several novel tasks are presented, including business taxonomy construction, the rationale of trading, and contract disambiguation. Some new models are constructed for sales prediction, stock market prediction, and sentiment analysis. Tools and corpora related to FinTech and NLP are demonstrated. Furthermore, systems in the FinSBD shared task break the state-of-the-art of Sentence Boundary Detection in PDF Noisy Text in the Financial Domain.

We are immensely grateful to the Shared Task Organizers Sira Ferradans, Abderrahim Ait-Azzi, Guillaume Hubert, and Houda Bouamor. We also thank all of the Program Committee Members for their thorough review of the submissions. Besides, we would like to express our gratitude to MOST Joint Research Center of AI Technology and All Vista Healthcare for the financial support. Finally, many thanks to all participants for submitting their fine work and sharing their ideas. Without their efforts, this workshop could not be successful.

Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, Hsin-Hsi Chen

FinNLP 2019 Organizers

July 2019



# Committees

## Organizers

Chung-Chi Chen, National Taiwan University, Taiwan

Hen-Hsen Huang, National Chengchi University, Taiwan

Hiroya Takamura, National Institute of Advanced Industrial Science and Technology, Japan

Hsin-Hsi Chen, National Taiwan University, Taiwan

## Shared Task Organizers

Sira Ferradans, Fortia Financial Solutions, France

Abderrahim Ait-Azzi, Fortia Financial Solutions, France

Guillaume Hubert, Fortia Financial Solutions, France

Houda Bouamor, Carnegie Mellon University, Qatar

## Program Committee

Paulo Alves, Universidade Católica Portuguesa, Portugal

Avi Arampatzis, Democritus University of Thrace, Greece

Alexandra Balahur, European Commission's Joint Research Centre, Italy

Houda Bouamor, Carnegie Mellon University, Qatar

Paul Buitelaar, National University of Ireland Galway, Ireland

Damir Cavar, Indiana University, USA

Sunandan Chakraborty, Indiana University, USA

Brian Davis, Maynooth University, Ireland

Sira Ferradans, Fortia Financial Solutions, France

André Freitas, The University of Manchester, UK

Els Lefever, Ghent University, Belgium

Sheng Li, Google, USA

Nedim Lipka, Adobe Inc., USA

Heiner Stuckenschmidt, University of Mannheim, German

Ming-Feng Tsai, National Chengchi University, Taiwan

Chuan-Ju Wang, Academia Sinica, Taiwan

Wlodek Zadrozny, University of North Carolina in Charlotte, USA

Manel Zarrouk, National University of Ireland Galway, Ireland



# Table of Contents

<b>Business Taxonomy Construction Using Concept-Level Hierarchical Clustering</b> Haodong Bai, Frank Xing, Erik Cambria and Win-Bin Huang.....	1
<b>Towards Disambiguating Contracts for their Successful Execution - A Case from Finance Domain</b> Preethu Rose Anish, Abhishek Sainani, Nitin Ramrakhiyani, Sachin Pawar, Girish K Palshikar and Smita Ghaisas.....	8
<b>Rationale Classification for Educational Trading Platforms</b> Annie Ying and Pablo Duboue.....	14
<b>CoFiF: A Corpus of Financial Reports in French Language</b> Tobias Daudert and Sina Ahmadi.....	21
<b>Step-wise Refinement Classification Approach for Enterprise Legal Litigation</b> Ying Mao, Xian Wang, Jianbo Tang and Changliang Li.....	27
<b>CoSACT: A Collaborative Tool for Fine-Grained Sentiment Annotation and Consolidation of Text</b> Tobias Daudert, Manel Zarrouk and Brian Davis.....	34
<b>Financial Text Data Analytics Framework for Business Confidence Indices and Inter Industry Relations</b> Hiroki Sakaji, Ryota Kuramoto, Hiroyasu Matsushima, Kiyoshi Izumi, Takashi Shimada and Keita Sunakawa.....	40
<b>Learning to Learn Sales Prediction with Social Media Sentiment</b> Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, Zihan Liu, Yan Xu, Cong Gao and Pascale Fung.....	47
<b>Leveraging BERT to Improve the FEARS Index for Stock Forecasting</b> Linyi Yang, Ruihai Dong, Tin Lok James Ng and Yang Xu.....	54
<b>Economic Causal-Chain Search Using Text Mining Technology</b> Kiyoshi Izumi and Hiroki Sakaji.....	61
<b>Transformer-Based Capsule Network for Stock Movement Prediction</b> Jintao Liu, Hongfei Lin, Xikai Liu, Bo Xu, Yuqi Ren, Yufeng Diao and Liang Yang.....	66
<b>The FinSBD-2019 Shared Task: Sentence Boundary Detection in PDF Noisy Text in the Financial Domain</b> Abderrahim Ait Azzi, Houda Bouamor and Sira Ferradans.....	74

<b>AIG Investments.AI at the FinSBD Task: Sentence Boundary Detection through Sequence Labelling and BERT Fine-tuning</b>	
Jinhua Du, Yan Huang and Karo Moilanen.....	81
<b>aiai at FinSBD Task: Sentence Boundary Detection in Noisy Texts from Financial Documents Using Deep Attention Model</b>	
Ke Tian and Zi Jun Peng.....	88
<b>Pluto: A Deep Learning Based Watchdog for Anti Money Laundering</b>	
Hao-Yuan Chen, Shang-Xuan Zou and Cheng-Lung Sung.....	93
<b>From Creditworthiness to Trustworthiness with Alternative NLP/NLU Approaches</b>	
Charles Crouspeyre, Eleonore Alesi and Karine Lespinasse.....	96
<b>On a Chatbot Conducting a Virtual Dialogue in Financial Domain</b>	
Boris Galitsky and Dmitry Ilvovsky.....	99
<b>mhirano at the FinSBD Task: Pointwise Prediction Based on Multi-layer Perceptron for Sentence Boundary Detection</b>	
Masanori Hirano, Hiroki Sakaji, Kiyoshi Izumi and Hiroyasu Matsushima.....	102
<b>NUIG at the FinSBD Task: Sentence Boundary Detection for Noisy Financial PDFs in English and French</b>	
Tobias Daudert and Sina Ahmadi.....	108
<b>Sentence Boundary Detection of Financial Data with Domain Knowledge Enhancement and Bilingual Training</b>	
Mingyu Wan, Rong Xiang, Emmanuele Chersoni, Natalia Klyueva, Kathleen Ahrens, Bin Miao, David Broadstock, Jian Kang, Amos Yung and Chu-Ren Huang.....	115
<b>HITS-SBD at the FinSBD Task: Machine Learning vs. Rule-based Sentence Boundary Detection</b>	
Mehwish Fatima and Mark-Christoph Mueller.....	122
<b>AI_Blues at FinSBD Shared Task: CRF-based Sentence Boundary Detection in PDF Noisy Text in the Financial Domain</b>	
Ditty Mathew and Chinnappa Guggilla.....	130



## Conference Program

- 9:00-9:15 Opening Remarks
- 9:15-9:40 Business Taxonomy Construction Using Concept-Level Hierarchical Clustering  
Haodong Bai, Frank Xing, Erik Cambria and Win-Bin Huang
- 9:40-10:05 Towards Disambiguating Contracts for their Successful Execution - A Case from Finance Domain  
Preethu Rose Anish, Abhishek Sainani, Nitin Ramrakhiyani, Sachin Pawar, Girish K Palshikar and Smita Ghaisas
- 10:05-10:30 Rationale Classification for Educational Trading Platforms  
Annie Ying and Pablo Duboue
- 10:30-11:00 Break
- 11:00-11:25 CoFiF: A Corpus of Financial Reports in French Language  
Tobias Daudert and Sina Ahmadi
- 11:25-11:50 Step-wise Refinement Classification Approach for Enterprise Legal Litigation  
Ying Mao, Xian Wang, Jianbo Tang and Changliang Li
- 11:50-12:15 CoSACT: A Collaborative Tool for Fine-Grained Sentiment Annotation and Consolidation of Text  
Tobias Daudert, Manel Zarrouk and Brian Davis
- 12:15-13:30 Lunch
- 13:30-13:55 Financial Text Data Analytics Framework for Business Confidence Indices and Inter-Industry Relations  
Hiroki Sakaji, Ryota Kuramoto, Hiroyasu Matsushima, Kiyoshi Izumi, Takashi Shimada and Keita Sunakawa
- 13:55-14:20 Learning to Learn Sales Prediction with Social Media Sentiment  
Zhaojiang Lin, Andrea Madotto, Genta Indra Winata, Zihan Liu, Yan Xu, Cong Gao and Pascale Fung
- 14:20-14:45 Leveraging BERT to Improve the FEARS Index for Stock Forecasting  
Linyi Yang, Ruihai Dong, Tin Lok James Ng and Yang Xu
- 14:45-15:10 Economic Causal-Chain Search Using Text Mining Technology  
Kiyoshi Izumi and Hiroki Sakaji
- 15:10-15:30 Break
- 15:30-15:55 Transformer-Based Capsule Network for Stock Movement Prediction  
Jintao Liu, Hongfei Lin, Xikai Liu, Bo Xu, Yuqi Ren, Yufeng Diao and Liang Yang

- 15:55-16:15 **The FinSBD-2019 Shared Task: Sentence Boundary Detection in PDF Noisy Text in the Financial Domain**  
Abderrahim Ait Azzi, Houda Bouamor and Sira Ferradans
- 16:15-16:25 **AIG Investments.AI at the FinSBD Task: Sentence Boundary Detection through Sequence Labelling and BERT Fine-tuning**  
Jinhua Du, Yan Huang and Karo Moilanen
- 16:25-16:35 **aiai at FinSBD Task: Sentence Boundary Detection in Noisy Texts from Financial Documents Using Deep Attention Model**  
Ke Tian and Zi Jun Peng
- 16:40-18:00 **Poster session**
- Pluto: A Deep Learning Based Watchdog for Anti Money Laundering**  
Hao-Yuan Chen, Shang-Xuan Zou and Cheng-Lung Sung
- From Creditworthiness to Trustworthiness with Alternative NLP/NLU Approaches**  
Charles Crouspeyre, Eleonore Alesi and Karine Lespinasse
- On a Chatbot Conducting a Virtual Dialogue in Financial Domain**  
Boris Galitsky and Dmitry Ilvovsky
- mhirano at the FinSBD Task: Pointwise Prediction Based on Multi-layer Perceptron for Sentence Boundary Detection**  
Masanori Hirano, Hiroki Sakaji, Kiyoshi Izumi and Hiroyasu Matsushima
- NUIG at the FinSBD Task: Sentence Boundary Detection for Noisy Financial PDFs in English and French**  
Tobias Daudert and Sina Ahmadi
- Sentence Boundary Detection of Financial Data with Domain Knowledge Enhancement and Bilingual Training**  
Mingyu Wan, Rong Xiang, Emmanuele Chersoni, Natalia Klyueva, Kathleen Ahrens, Bin Miao, David Broadstock, Jian Kang, Amos Yung and Chu-Ren Huang
- HITS-SBD at the FinSBD Task: Machine Learning vs. Rule-based Sentence Boundary Detection**  
Mehwish Fatima and Mark-Christoph Mueller
- AI\_Blues at FinSBD Shared Task: CRF-based Sentence Boundary Detection in PDF Noisy Text in the Financial Domain**  
Ditty Mathew and Chinnappa Guggilla