

EMNLP 2021

**ECONLP — Economics and Natural Language Processing**

**Proceedings of the Third Workshop on Economics and  
Natural Language Processing (ECONLP 2021)**

November 11, 2021  
Punta Cana, Dominican Republic and Online

©2021 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 978-1-954085-84-8

## Introduction

Welcome to ECONLP 2021! After launching the first Workshop on Economics and Natural Language Processing (ECONLP) at ACL 2018 in Melbourne, Australia and running the second one at EMNLP-IJCNLP 2019 in Hong Kong, China, the third ECONLP workshop is held at EMNLP 2021 in a hybrid manner, physically in Punta Cana, Dominican Republic, and virtually as an online event.

This workshop reflects the increasing relevance of natural language processing (NLP) for regional, national and international economy, both in terms of already operational language technology products and systems, as well as newly emerging methodologies and techniques as a response to new requirements at the disciplinary intersection of economics and NLP. The focus of the workshop is on the many ways how NLP supports and influences business relations and procedures, economic transactions, and the roles of human and computational actors involved in commercial activities (such as e-commerce).

The main topics ECONLP addresses include (definitely not excluding other areas of relevance for the broad workshop theme):

- NLP-based (stock) *market analytics*, e.g., prediction of economic performance indicators (trend prediction, performance forecasting, etc.), by analyzing verbal statements of enterprises, businesses, companies, and associated institutional, legal or administrative actors
- NLP-based *product analytics*, e.g., based on social and traditional mass media by monitoring, summarizing reviews, classifying and mining comments or complaint messages and other (non-) verbal types of customer reactions to products or services
- NLP-based *customer analytics*, e.g., client profiling, tracking product/company preferences, screening customer reviews or complaints, identifying high-influentials (peers) in economy-related communication networks
- NLP-based *organization/enterprise analytics*, e.g., tracing and pro-actively altering social images of organizational actors, risk prediction, fraud analysis, predictive analysis of annual business, sustainability and auditing reports
- NLP-based *analysis of macro-economic phenomena* in which national economies and the (inter)national banking system (IMF, Fed, PBoC, ECB) play an influential role
- Analysis of *market sentiments and emotions* as evident from consumers' and enterprises' verbal behavior traces and their communication strategies about products, services or market performance
- *Competitive intelligence* services based on NLP tooling
- Relationships and interactions between *quantitative (structured) economic data* (e.g., those available from sales databases and associated time series data, log data) and *qualitative (unstructured verbal) economic data* (press releases, newswire streams, social media contents, etc.)
- *Organizational information management* based on the content-based assembling, packaging and archiving of verbal communication streams in organizations/enterprises (emails, meeting minutes, business letters, internal reporting, etc.)
- *Credibility and trust models* for business agents involved in the economic process (e.g., as traders, sellers, advertisers) extracted from text/opinion mining their communication behavior (including historic or legacy data)

- *Deception or fake information recognition* related to economic objects (such as products, advertisements, etc.) or economic actors (such as industries, companies, etc.), including opinion spam targeting or emanating from economic actors and processes
- Verbally fluent *software agents* (chat bots for counseling, sales and marketing) as reliable actors in economic processes serving business interests, e.g., embodying models of persuasion, information biases, fair trading
- *Enterprise search engines* (e-commerce, e-marketing) involving NLP analytics
- *Consumer search engines*, market monitors, product/service recommender systems involving NLP analytics
- *Client-supplier interaction platforms* (e.g., portals, helps desks, newsgroups) and transaction support systems based on written or spoken natural language communication
- *Multi-media and multi-modality interaction platforms*, including written/spoken language channels, triggering or supporting economic processes
- Specialized modes of *information extraction and text mining* in economic domains, e.g., temporal event or transaction mining
- *Information aggregation* from economy-related single sources (e.g., review summaries, automatic threading)
- *Text generation* in economic domains, e.g., review generation, complaint response generation
- *Ontologies and knowledge graphs* for economics and adaptation of general-domain *lexicons* for economic NLP
- *Corpora and annotations policies* (guidelines, metadata schemata, etc.) for economics-related NLP
- *Economy-specific text genres* (business reports, sustainability reports, auditing documents, product reviews, economic newswire, business letters, law documents, etc.) and their usage for NLP
- Dedicated *software resources* for economic NLP (e.g., NER taggers, sublanguage parsers, pipelines or domain-adapted end-to-end systems for processing economic discourse)

Two types of papers were solicited for the ECONLP 2021 workshop:

- Long papers (8+1 pages) should describe solid results with strong experimental, empirical or theoretical/formal backing,
- Short papers (4+1 pages) should describe work in progress where interesting novel, yet still preliminary results have been worked out.

We received this year 16 submissions and based on a rigorous review process, we accepted 5 of them as long papers, 7 as short papers and rejected 4 submissions. Accordingly, the acceptance rate was 75%. The OC of ECONLP 2021 decided on such a proportionally high number of acceptances due to the exceptional quality of the papers under review. The acceptance/rejection ratio amounts to 3.0.

After three rounds, ECONLP has become ACL's anchor workshop for foundational and applied research at the intersection of Economics and Natural Language Processing. Submission numbers stabilize in the range of 15 to 20 papers, with acceptance ratios in the interval of 50 up to 75%.

Yet, ECONLP is only part of a series of workshops at the intersection of NLP, economics, and related scientific fields, such as artificial intelligence, machine learning, data science, knowledge discovery, Web science, and information retrieval. Two companion workshops are closely affiliated with other NLP conferences, namely this year's editions of

- *ECNLP 4 — 4th Workshop on e-Commerce and NLP @ ACL 2021* (with previous editions taking place at the Web Conference (WWW 2019 and 2020) and ACL 2020)
- *FNP 2021 — 3rd Financial Narrative Processing Workshop @ University of Lancaster, UK* (with previous editions taking place at COLING 2020, NoDaLiDa '19 (jointly with the MultiLing Financial Summarisation (FNP-FNS 2020) workshop), and LREC 2018)

But the theme of this workshop has attracted an even wider range of neighboring scientific communities, including NeurIPS, IJCAI, KDD, and SIGIR. Without aiming for completeness, the list below enumerates some recent events with a similar thematic scope:

- *FAIF 2020 — (1st) Workshop on Fair AI in Finance @ NeurIPS 2020*
- *FinNLP & FinSim & FinSDB 2020 — 2nd Workshop on Financial Technology and Natural Language Processing (FinNLP) with FinSim and FinSDB-2 Shared Task @ IJCAI-PRICAI 2020*
- *KDD-MLF 2021 — 4th Workshop on Machine Learning in Finance @ KDD 2021*
- *eCom 2021 — 5th ACM SIGIR Workshop on eCommerce @ SIGIR 2021*

These manifold activities are exciting to witness and, at the same time, indicate the added value of approaches complementing each other in this truly interdisciplinary field.

Finally, we want to thank all our collaborators. First of all, our colleagues who submitted their best work to this workshop. Their efforts are crucial for the development of this rapidly evolving field. Second, we are indebted to our PC members whose thorough, balanced and timely reviews were the basis for properly selecting the best-quality papers presented at this workshop. Last but not least, our sincere thanks go to our keynote speaker, Gerard Hoberg, from the Marshall School of Business, University of Southern California, Los Angeles, CA, USA, for his valuable contribution to this year's ECONLP. Finally, we hope the attendants of the workshop, whether physically or virtually present, will enjoy the presentations and discussions in Punta Cana or online.

The organizers of ECONLP 2021

Udo Hahn  
Véronique Hoste  
Amanda Stent



**Organizers:**

Udo Hahn	Friedrich-Schiller-Universität Jena, Jena, Germany (Chairman)
Véronique Hoste	Ghent University, Ghent, Belgium
Amanda Stent	Bloomberg, New York City, NY, USA

**Program Committee:**

Sven Buechel	Friedrich-Schiller-Universität Jena, Jena, Germany
David Carmel	Amazon, Haifa, Israel
Michael Chau	School of Business, University of Hong Kong, Hong Kong, China
Paulo Cortez	University of Minho, Guimarães, Portugal
Sanjiv Ranjan Das	Santa Clara University, Santa Clara, CA, USA
Brian Davis	School of Computing, Dublin City University, Dublin, Ireland
Luciano Del Corro	Goldman Sachs, Frankfurt/M., Germany
Lipika Dey	Tata Consultancy Services (TCS) Innovation Lab, New Delhi, India
Giuseppe Di Fabbri	VUI, Inc., Boston, MA, USA & Trento, Italy
Flavius Frasin	Erasmus University, Rotterdam, Netherlands
Anjan Goswami	Adobe Inc., San Francisco, CA, USA
Petr Hájek	University of Pardubice, Pardubice, Czech Republic
Yulan He	University of Warwick, Coventry, UK
Qing Li	Southwestern University of Finance and Economics, Sichuan, Chengdu, China
Xiaodong Li	Hohai University, Nanjing, China
Pekka Malo	Aalto University, Aalto, Finland
Igor Mozetič	Jožef Stefan Institute, Ljubljana, Slovenia
Viktor Pekar	Aston University, Birmingham, UK
Nicolas Pröllochs	Universität Giessen, Giessen, Germany
Samuel Rönqvist	University of Turku, Turku, Finland
Hiroki Sakaji	University of Tokyo, Tokyo, Japan
Kazuhiro Seki	University of Kobe, Kobe, Japan
Sameena Shah	JPMorgan Chase, New York, NY, USA
Kiyooki Shirai	Japan Advanced Institute of Science and Technology (JAIST), Nomi, Japan
Heiner Stuckenschmidt	Universität Mannheim, Mannheim, Germany
Jacopo Tagliabue	Coveo Labs, New York City, NY, USA
Mengting Wan	Microsoft, Redmond, WA, USA
Frank Z. Xing	Nanyang Technological University (NTU), Singapore, Singapore
Wlodek W. Zadrozny	University of North Carolina at Charlotte, Charlotte, NC, USA
Zhu (Drew) Zhang	Iowa State University, Ames, IA, USA

**Invited Speaker:**

Gerard Hoberg	Marshall School of Business, University of Southern California, Los Angeles, CA, USA
---------------	--

# **Keynote talk: Using NLP to Model Product Market Boundaries, Market Structure, and Innovation**

**Gerard Hoberg**

Marshall School of Business, University of Southern California, Los Angeles, CA, USA

## **Abstract**

Research in financial economics utilizing natural language processing has modeled how U.S. companies compete across product markets, how they innovate, and how supply chains are organized. This presentation will summarize these models, how they have evolved to the present day, and new work in progress. A key technical foundation is that these concepts of market structure, which were developed by economic theorists, have relatedness and overlap as their very intellectual foundation. Hence NLP methods (new and old) that focus on textual relatedness make up an ideal toolkit for analyzing these economic concepts, which are also central to many applications aimed at understanding and predicting business success.

One of the most compelling properties of an NLP approach is that market structures can be updated and redrawn every year at low cost as textual corpora are updated. This observation indicates the use of NLP to measure innovation. This has led to many insights on how companies manage growth, whether and when they go public, and how they manage the risk of disruption. This work continues to evolve with a focus on innovation, disaggregating complex product portfolios, and to build NLP measures for private firms through their websites. Among other economic findings, this work reveals that U.S. companies are becoming increasingly complex and broad in their product offerings, a result that is not seen using standard numerical databases in the public domain. Work using these same methods has also been deployed in related settings such as modeling systemic risk.



## Table of Contents

<i>A Fine-Grained Annotated Corpus for Target-Based Opinion Analysis of Economic and Financial Narratives</i>	
Jiahui Hu and Patrick Paroubek .....	1
<i>EDGAR-CORPUS: Billions of Tokens Make The World Go Round</i>	
Lefteris Loukas, Manos Fergadiotis, Ion Androutsopoulos and Prodromos Malakasiotis .....	13
<i>The Global Banking Standards QA Dataset (GBS-QA)</i>	
Kyunghwan Sohn, Sunjae Kwon and Jaesik Choi .....	19
<i>Corporate Bankruptcy Prediction with BERT Model</i>	
Alex Kim and Sangwon Yoon .....	26
<i>Is Domain Adaptation Worth Your Investment? Comparing BERT and FinBERT on Financial Tasks</i>	
Bo Peng, Emmanuele Chersoni, Yu-Yin Hsu and Chu-Ren Huang .....	37
<i>From Stock Prediction to Financial Relevance: Repurposing Attention Weights to Assess News Relevance Without Manual Annotations</i>	
Luciano Del Corro and Johannes Hoffart .....	45
<i>Privacy enabled Financial Text Classification using Differential Privacy and Federated Learning</i>	
Priyam Basu, Tiasa Singha Roy, Rakshit Naidu and Zumrut Muftuoglu .....	50
<i>Using Word Embedding to Reveal Monetary Policy Explanation Changes</i>	
Akira Matsui, Xiang Ren and Emilio Ferrara .....	56
<i>To What Extent Can English-as-a-Second Language Learners Read Economic News Texts?</i>	
Yo Ehara .....	62
<i>Effective Use of Graph Convolution Network and Contextual Sub-Tree for Commodity News Event Extraction</i>	
Meisin Lee, Lay-Ki Soon and Eu-Gene Siew .....	69
<i>Cryptocurrency Day Trading and Framing Prediction in Microblog Discourse</i>	
Anna Paula Pawlicka Maule and Kristen Johnson .....	82
<i>Extracting Economic Signals from Central Bank Speeches</i>	
Maximilian Ahrens and Michael McMahon .....	93



# Conference Program

**Thursday, November 11, 2021**

09:00–09:20 *Opening Remarks and Status Report on Economics and NLP*  
Udo Hahn

**09:20–10:30 Session 1: Datasets for Economic NLP**

9:20–9:50 *A Fine-Grained Annotated Corpus for Target-Based Opinion Analysis of Economic and Financial Narratives*  
Jiahui Hu and Patrick Paroubek

9:50–10:10 *EDGAR-CORPUS: Billions of Tokens Make The World Go Round*  
Lefteris Loukas, Manos Fergadiotis, Ion Androutsopoulos and Prodromos Malakasiotis

10:10–10:30 *The Global Banking Standards QA Dataset (GBS-QA)*  
Kyunghwan Sohn, Sunjae Kwon and Jaesik Choi

**10:30–11:00 Coffee Break**

**11:00–12:10 Session 2: Transformer-based Methodologies in Economic NLP**

11:00–11:30 *Corporate Bankruptcy Prediction with BERT Model*  
Alex Kim and Sangwon Yoon

11:30–11:50 *Is Domain Adaptation Worth Your Investment? Comparing BERT and FinBERT on Financial Tasks*  
Bo Peng, Emmanuele Chersoni, Yu-Yin Hsu and Chu-Ren Huang

11:50–12:10 *From Stock Prediction to Financial Relevance: Repurposing Attention Weights to Assess News Relevance Without Manual Annotations*  
Luciano Del Corro and Johannes Hoffart

**Thursday, November 11, 2021 (continued)**

**12:10–12:45 Lunch Break**

12:45–13:30 *Keynote: Using NLP to Model Product Market Boundaries, Market Structure, and Innovation*  
Gerard Hoberg

**13:30–14:30 Session 3: Applications in Economic NLP (1)**

13:30–13:50 *Privacy enabled Financial Text Classification using Differential Privacy and Federated Learning*  
Priyam Basu, Tiasa Singha Roy, Rakshit Naidu and Zumrut Muftuoglu

13:50–14:10 *Using Word Embedding to Reveal Monetary Policy Explanation Changes*  
Akira Matsui, Xiang Ren and Emilio Ferrara

14:10–14:30 *To What Extent Can English-as-a-Second Language Learners Read Economic News Texts?*  
Yo Ehara

**14:30–14:45 Short Break**

**14:45–16:15 Session 4: Applications in Economic NLP (2)**

14:45–15:15 *Effective Use of Graph Convolution Network and Contextual Sub-Tree for Commodity News Event Extraction*  
Meisin Lee, Lay-Ki Soon and Eu-Gen Siew

15:15–15:45 *Cryptocurrency Day Trading and Framing Prediction in Microblog Discourse*  
Anna Paula Pawlicka Maule and Kristen Johnson

15:45–16:15 *Extracting Economic Signals from Central Bank Speeches*  
Maximilian Ahrens and Michael McMahon

**Thursday, November 11, 2021 (continued)**

**16:15–16:45** Coffee Break

**16:45–17:30** *Discussion Round: Planning for an ECONLP Challenge Competition*

**17:30–17:45** *Concluding Remarks*

