

SHUBHANSHU MISHRA

mishra@shubhanshu.com · shubhanshu.com

EDUCATION

University of Illinois at Urbana-Champaign

Doctor of Philosophy (Ph.D.) Library and Information Science

Aug 2013 - May 2020

Thesis: Information extraction from digital social trace data with applications to social media and scholarly communication data

Indian Institute of Technology Kharagpur

Bachelors and Masters in Science Mathematics and Computing

Jul 2007 - May 2012

Thesis: Analysis of Social Media Data to determine Positive and Negative Influential Nodes in the Network

EXPERIENCE

Twitter, Inc., United States of America

Senior Machine Learning Researcher, Content Understanding Research

Jan 2022 - Present

Machine Learning Researcher, Content Understanding Research

Aug 2019 - Dec 2021

Intern, Content Understanding & Applied Deep Learning

Jul 2018 - Sep 2018

Citrix, Bangalore, India

Software Engineer, NetScaler INFRA Team

Jul 2012 - Jul 2013

Barclays Capital, Singapore

Global Technology Analyst, Commodities

May 2011 - Jul 2011

Global Venture Lab, Finland

Lead Web Developer

Dec 2009 - Jan 2010

National University of Singapore, Singapore

Research Assistant at Institute of Systems Science

May 2009 - Jul 2009

ACADEMIC HONORS & AWARDS

Impact Recognition Award - CSCW

Oct 2021

Best Poster Award - UIUC Student Poster Session

Mar 2020

Best student paper award - ASIST SIGMET Workshop

Nov 2018

Graduate Teacher Certificate

May 2018

University of Illinois GIS Day Runner-up (Research Quality)

Nov 2017

Kishore Vaigyanik Protsahan Yojana Scholar

2007-2012

3rd rank in Regional Mathematics Olympiad, Uttar Pradesh, India

Dec 2006

TEACHING

Tutorial presenter, Multiple venues

Sep 2019 - Current

Tutorial on hands on advanced machine learning for information extraction from tweets tasks, data, and open source tools. Details at: <https://socialmediaie.github.io/tutorials/>

Co-instructor - Network Analysis

Spring 2018

Teaching network analysis applications on social media data using python.

Teaching Assistant - Network Analysis,

Summer 2017

As part of the Information Science and Engineering Summer Program organized for Chinese students, sponsored by Global Education and Training (GET)

Teaching Assistant - Foundations of Information Processing

Spring 2017

Co-instructor - Data Mining Applications

Fall 2016

Listed in Teachers Ranked as Excellent By Their Students!

PROJECTS

Multilingual Language Model Pretraining via Translation Pair Prediction

2021

LMSOC: An Approach for Socially Sensitive Pretraining

2021

Image Crop Analysis

2021

Social Media Information Extraction

2019

WikiCSSH - Computer Science Subject Headings from Wikipedia

2020

Profiling authors and articles based on novelty, expertise and self-citation

2016

ConText - Tool for extracting and analyzing network data from text

2013

Entity Extractor using CRF and Deep Learning

2014

SAIL - Sentiment Analysis and Incremental Learning

2014

SentiNets - Identification of enthusiastic supporters of social causes using text and network analysis

2013

ReadLater - Chrome extension with 10,000+ users with 4.5/5 rating

2012

- Kulkarni, V., Mishra, S., & Haghighi, A. (2021). {LMSOC}: An Approach for Socially Sensitive Pretraining. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2967–2975.
- Mishra, S., & Haghighi, A. (2021). Improved Multilingual Language Model Pretraining for Social Media Text via Translation Pair Prediction. *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, 381–388.
- Mishra, S., Prasad, S., & Mishra, S. (2021). Exploring Multi-Task Multi-Lingual Learning of Transformer Models for Hate Speech and Offensive Speech Identification in Social Media. *SN Computer Science*, 2(2), 72. <https://doi.org/10.1007/s42979-021-00455-5>
- Yee, K., Tantipongpipat, U., & Mishra, S. (2021). Image Cropping on Twitter: Fairness Metrics, their Limitations, and the Importance of Representation, Design, and Agency. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–24. <https://doi.org/10.1145/3479594>
- Han, K., Yang, P., Mishra, S., & Diesner, J. (2020). WikiCSSH: Extracting Computer Science Subject Headings from Wikipedia. *Workshop on Scientific Knowledge Graphs (SKG 2020)*.
- Mishra, S. (2020a). Information Extraction from Digital Social Trace Data with Applications to Social Media and Scholarly Communication Data. *ACM SIGIR Forum*, 54(1).
- Mishra, S. (2020b). Non-neural Structured Prediction for Event Detection from News in Indian Languages. In P. Mehta, T. Mandl, P. Majumder, & M. Mitra (Eds.), *Working notes of fire 2020 - forum for information retrieval evaluation*. CEUR Workshop Proceedings, CEUR-WS.org.
- Mishra, S., & Collier, D. (2020). A Framework for Generating Annotated Social Media Corpora with Demographics, Stance, Civility, and Topicality. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3757554>
- Mishra, S., & Mishra, S. (2020a). Scubed at 3C task A - A simple baseline for citation context purpose classification. *Proceedings of the 8th International Workshop on Mining Scientific Publications*, 59–64.
- Mishra, S., & Mishra, S. (2020b). Scubed at 3C task B - A simple baseline for citation context influence classification. *Proceedings of the 8th International Workshop on Mining Scientific Publications*, 65–70.
- Mishra, S., Prasad, S., & Mishra, S. (2020). Multilingual Joint Fine-tuning of Transformer models for identifying Trolling, Aggression and Cyberbullying at TRAC 2020. *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, 120–125.
- Parulian, N. N., Lu, T., Mishra, S., Avram, M., & Diesner, J. (2020). Effectiveness of the Execution and Prevention of Metric-Based Adversarial Attacks on Social Network Data †. *Information*, 11(6), 306. <https://doi.org/10.3390/info11060306>
- Avram, M. V., Mishra, S., Parulian, N. N., & Diesner, J. (2019). Adversarial perturbations to manipulate the perception of power and influence in networks. *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 986–993.
- Collier, D., Mishra, S., Houston, D., Hensley, B., Mitchell, S., & Hartlep, N. (2019). Who is Most Likely to Oppose Federal Tuition-Free College Policies? Investigating Variable Interactions of Sentiments to America’s College Promise. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3423054>
- Collier, D. A., Mishra, S., Houston, D. A., Hensley, B. O., & Hartlep, N. D. (2019). Americans ‘support’ the idea of tuition-free college: an exploration of sentiment and political identity signals otherwise. *Journal of Further and Higher Education*, 43(3), 347–362. <https://doi.org/10.1080/0309877X.2017.1361516>
- Mishra, S. (2019). Multi-dataset-multi-task Neural Sequence Tagging for Information Extraction from Tweets. *Proceedings of the 30th ACM Conference on Hypertext and Social Media - HT ’19*, 283–284. <https://doi.org/10.1145/3342220.3344929>
- Mishra, S., & Diesner, J. (2019). Capturing Signals of Enthusiasm and Support Towards Social Issues from Twitter. *Proceedings of the 5th International Workshop on Social Media World Sensors - SIdEWayS’19*, 19–24. <https://doi.org/10.1145/3345645.3351104>
- Mishra, S., & Mishra, S. (2019). 3Idiots at HASOC 2019: Fine-tuning Transformer Neural Networks for Hate Speech Identification in Indo-European Languages. *Proceedings of the 11th annual meeting of the Forum for Information Retrieval Evaluation*, 208–213.
- Mishra, S., & Diesner, J. (2018). Detecting the Correlation between Sentiment and User-level as well as Text-Level Meta-data from Benchmark Corpora. *Proceedings of the 29th on Hypertext and Social Media - HT ’18*, 2–10. <https://doi.org/10.1145/3209542.3209562>
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018a). Expertise as an aspect of author contributions. *Metrics 2018: Workshop on Informetric and Scientometric Research (SIG/MET)*.
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018b). Self-citation is the hallmark of productive authors, of any gender (N. O. Schiller, Ed.). *PLoS ONE*, 13(9), e0195773. <https://doi.org/10.1371/journal.pone.0195773>
- Addawood, A., Rezapour, R., Mishra, S., Schneider, J., & Diesner, J. (2017). Developing an Information Source Lexicon. *Prioritising Online Content workshop co-located at NIPS*.

- Mishra, S. (2017). SCTG: Social Communications Temporal Graph – A novel approach to visualize temporal communication graphs from social data. *UIUC Data Science Day*.
- Mishra, S., & Diesner, J. (2016). Semi-supervised Named Entity Recognition in noisy-text. *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, 203–212.
- Mishra, S., & Torvik, V. I. (2016). Quantifying Conceptual Novelty in the Biomedical Literature. *D-Lib magazine : the magazine of the Digital Library Forum*, 22(9-10).
<https://doi.org/10.1045/september2016-mishra>
- Mishra, S., Diesner, J., Byrne, J., & Surbeck, E. (2015). Sentiment Analysis with Incremental Human-in-the-Loop Learning and Lexical Resource Customization. *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, 323–325. <https://doi.org/10.1145/2700171.2791022>
- Mishra, S., Agarwal, S., Guo, J., Phelps, K., Picco, J., & Diesner, J. (2014). Enthusiasm and support: alternative sentiment classification for social movements on social media. *Proceedings of the 2014 ACM conference on Web science - WebSci '14*, 261–262. <https://doi.org/10.1145/2615569.2615667>