

Sudipta Kar

✉ sudipta.kar.8080@gmail.com
🌐 sudiptakar.info ☎ 832.643.7277
🏠 Seattle, WA

Research Interests Natural Language Processing, Deep Learning, Machine Learning

Experience

- ❖ **Amazon** Applied Scientist at Alexa AI Seattle, WA July 2020 - Present
 - ◆ Proactive follow-up generation in dialogs.
- ❖ **Amazon** Applied Scientist Intern at Alexa Shopping NLU Seattle, WA Jan 2020 - May 2020
 - ◆ Continual Learning for Natural Language Understanding.
- ❖ **GE Healthcare** AI/ML Research Intern San Ramon, CA May 2019 - Aug 2019
 - ◆ Multi-modal information extraction from clinical data.
- ❖ **Intel AI** Deep Learning Engineering Intern San Diego, CA May 2018 - Aug 2018
 - ◆ Sentiment analysis of short texts and aspect based sentiment analysis using deep learning.
- ❖ **University of Houston** Instructional Assistant/ Research Assistant Aug 2015 - Jan 2020
 - ◆ Research on high-level story analysis and characterization.
 - ◆ Stylistic analysis of literary texts (books). Collaboration with Booxby.com.
 - ◆ Finding relevant questions and answers in community question answering systems.
 - ◆ Sentiment Analysis on Financial data.
 - ◆ TA : Introduction to Automata and Computability, Software Engineering.
- ❖ **CRTC, CSE-SUST** Researcher and Developer Sylhet, Bangladesh Jul 2013 - Jul 2015
 - ◆ **Pipilika Top News** : Automatic breaking and similar news detector. Live at <http://news.pipilika.com>
 - ◆ Basic NLP tools like POS tagger, keyword extractor, stemmer for Bangla.
 - ◆ Bangla Spell Cheking game **Shobdokolpodroom** .
- ❖ **Webcraft Bangladesh Ltd** Part-time Game Developer Jan 2012- Jun 2012
 - ◆ Cool Match Madness: Cross-platform mobile game for iOS, Android, and Nexus.

Education

- ❖ **University of Houston** Ph.D. in Computer Science Aug 2015 - May 2020
 - ◆ **Advisor** : Dr. Thamar Solorio
 - ◆ **Thesis**: Automatic Characterization of Stories.
- ❖ **Shahjalal University of Science and Technology** B.Sc. in CSE Jan 2009 - May 2013
 - ◆ **Thesis**: Clustering Bangla News Documents.

Publications

Google Scholar https://scholar.google.com/citations?user=MK_tdhEAAA&hl=en

1. Besnik Fetahu, **Sudipta Kar** , Zhiyu Chen, Oleg Rokhlenko, Shervin Malmasi. **SemEval-2023 Task 2: Fine-grained Multilingual Complex Named Entity Recognition (MultiCoNER 2)** . (ACL 2023).
2. Christopher Richardson, **Sudipta Kar** , Anjishnu Kumar, Anand Ramachandran, Omar Zia Khan, Zeynab Raeesy, Abhinav Sethy. **Learning to Retrieve Engaging Follow-Up Queries**. (EACL 2023).
3. **Sudipta Kar** , Giuseppe Castellucci, Simone Filice, Shervin Malmasi, Oleg Rokhlenko. **Preventing Catastrophic Forgetting in Continual Learning of New Natural Language Tasks**. (KDD 2022).
4. SUS. Chebolu, Paolo Rosso, **Sudipta Kar** , Thamar Solorio. **Survey on Aspect Category Detection** . (ACM Computing Surveys, 2022)
5. Shervin Malmasi, Anjie Fang*, Besnik Fetahu*, **Sudipta Kar** *, Oleg Rokhlenko. **MultiCoNER: a Large-scale Multilingual dataset for Complex Named Entity Recognition** . (COLING 2022).

6. Shervin Malmasi, Anjie Fang*, Besnik Fetahu*, **Sudipta Kar** *, Oleg Rokhlenko. **SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition (MultiCoNER)** . (NAACL 2022).
7. Henry Gorelick, BS Bijoy, SJ Saba, **Sudipta Kar** , MS Islam, MR Amin. **Syntax and Themes: How Context Free Grammar Rules and Semantic Word Association Influence Book Success**. (RANLP 2021).
8. Khondoker Ittehadul Islam, Md Saiful Islam, **Sudipta Kar** , Mohammad Ruhul Amin. **SentNoB: A Dataset for Analysing Sentiment on Noisy Bangla Texts**. (EMNLP 2021).
9. **Sudipta Kar** , Gustavo Aguilar, Mirella Lapata and Thamar Solorio. **Multi-view Story Characterization from Movie Plot Synopses and Reviews**. (EMNLP 2020).
10. Niloofar Safi Samghabadi, Afsheen Hatami, Mahsa Shafaei, **Sudipta Kar** and Thamar Solorio. **Attending the Emotions to Detect Online Abusive Language** (EMNLP 2020, Workshop on Online Abuse and Harms).
11. Parth Patwa, Gustavo Aguilar, **Sudipta Kar** , Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, Amitava Das. **Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets** (COLING, 2020).
12. Gustavo Aguilar, **Sudipta Kar** and Thamar Solorio. **LinCE: A Centralized Benchmark for Linguistic Code-switching Evaluation** (LREC, 2020).
13. Md. Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam and **Sudipta Kar** . **BanFakeNews: A Dataset for Detecting Fake News in Bangla** (LREC, 2020).
14. Mahsa Shafei, Niloofar Safi Samghabadi, **Sudipta Kar** and Thamar Solorio. **Age Suitability Rating: Predicting the MPAA Rating Based on Movie Dialogues** (LREC, 2020).
15. **Sudipta Kar** , Hsi-Ming Chang, Ratna Saripalli, Gopal B. Avinash. **Multimodal Identification of Medical Conditions from Chest X-ray Images and Radiology Reports**. (GE Data Science Symposium, Milwaukee, USA).
16. **Sudipta Kar** , Suraj Maharjan and Thamar Solorio. **Folksonomication: Predicting Tags for Movies from Plot Synopses using Emotion Flow encoded Neural Network**. (COLING 2018).
17. Niloofar Safi Samghabadi, Deepthi Mave, **Sudipta Kar** and Thamar Solorio. **RItual-UH at TRAC 2018 Shared Task: Aggression Identification**. (TRAC @ COLING 2018).
18. Suraj Maharjan, **Sudipta Kar** , Manuel Montes-y-Gómez, Fabio A. González, and Thamar Solorio. **Letting emotions flow: Success prediction by modeling the flow of emotions in books**. (NAACL 2018).
19. **Sudipta Kar** ; Suraj Maharjan; A. Pastor López-Monroy and Thamar Solorio. **MPST: A Corpus of Movie Plot Synopses with Tags**. (LREC 2018).
20. **Sudipta Kar** ; Suraj Maharjan; Thamar Solorio. **RItual-UH at SemEval-2017 Task 5: Sentiment Analysis on Financial Data Using Neural Networks**. (SemEval@ACL 2017).
21. Marc Franko-Salvador; **Sudipta Kar** ; Thamar Solorio; Paolo Rosso. **UH-PRHLT at SemEval-2016 Task 3: Combining Lexical and Semantic-based Features for Community Question Answering**. (SemEval@NAACL 2016).

Patent

1. **Book Analysis and Recommendation** US Patent US20200097538A1.

Professional Services

- **Co-organizer**: The Sixth workshop on Computational Approaches for Linguistic Code-switching (CALCS) at EMNLP 2023.
- **Co-organizer**: The First workshop on Bangla Language Processing (BLP) at EMNLP 2023.
- **Co-organizer**: The First NLP Hackathon in Bangladesh (36 hours, 150+ participants), January 2023.
- **Co-organizer**: SemEval 2022 Task 11: **MultiCoNER: Multilingual Complex Named Entity Recognition**.
- **Co-organizer**: SemEval 2020 Task 9: **SentiMix: Sentiment Analysis for Code-Mixed Social Media Text**.
- **Student Co-chair** of **NAACL' 2019 Student Research Workshop (SRW)** .
- **Area Chair/Reviewer**: ACL, AAAI, EMNLP, NAACL, EACL, LREC, SemEval, ACL SRW.

Invited Talks

1. Practical challenges in working with data, Fordham University, November 2020
2. 'Story of NLP' @ AI Geekfest, GE Healthcare, July 2019
3. Hot Topics in Department, NSM-UH, November 2018
4. Focus on Research, Dept. of Computer Science, UH, November 2018
5. Bengali.ai, October 2018
6. Guest lecture for COSC 6336 (Natural Language Processing), UH, April 2018

Technical Skills

Operating Systems : Mac OS X, Ubuntu, Windows

Languages : Python, Java, C/C++, R, PHP, Javascript, SQL, Lua, Scala, L^AT_EX.

Libraries and Tools : Transformers, Pytorch, Scikit-learn, Keras, NLTK, Weka, CoreNLP, Gensim, Jupyter, Pandas.

Media Coverage

- Featured in KHOU 11 News, December 2018.
https://www.youtube.com/watch?v=Ks9DxJ3t6_M
- Featured in Graduate Student Achievement, November 2018.
<http://www.uh.edu/nsm/news-events/stories/2018/1116-perfect-movie.php>
- Interviewed on Jamuna TV for works in Pipilika Bangla Search Engine, June 2014.
https://www.youtube.com/watch?v=2ynvf7_5DPc

Achievements and Extra Curricular Activities

- Created a podcast series in Bangla to provide guideline to new undergraduate students in Computer Science.
- Best PhD Student Award' 2019.
- Recipient of Cullen Travel Grant to attend LREC'2018.
- 1st place in SemEval'2016 Task 3 (Community Question Answering)
- 2nd place in SemEval'2017 Task 5 (Fine-Grained Sentiment Analysis on Financial Microblogs and News)
- 4th Place in National Collegiate Software Contest' 2011, Sylhet, Bangladesh
- Cultural Secretary of Bangladeshi Student Association at UH (2019-2020).
- Participated in National Collegiate Programming Contest' 2010, Sylhet, Bangladesh
- Solved around 150 problems in UVA Online Judge.
- Participated in CITI Financial IT Case Competition' 2011.
- Served as Elected General Secretary of CSE Society, SUST from November'2011 to February'2013.
- Worked as Executive Comity Member from May'2010 – May'2011 and Office Secretary from May'2011 – October'2012 in NONGAR (A Musical Organization in SUST)