C/O Swapan Kumar Das, Ichapur, Sasthibagan
P.O. Santragachi, Howrah
West Bengal, India, PIN:711104
✆ *(+91)-9831403301*
✉ *ayandas84@gmail.com*
✆ *http://cse.iitkgp.ac.in/ ayand/*

Ayan Das

Research Interests

- *Natural language processing:* NLP of low-resource languages, Cross-lingual NLP, NLP of Indian languages, Dependency parsing
- *Machine Learning:* Distributed machine learning, Application of deep learning architectures in NLP

Education

Jul'2013-Dec'2020	**Doctor of Philosophy (PhD)**, *Computer Science and Engineering,* Indian Institute of Technology Kharagpur.
Thesis:	***Effective methods for cross-lingual dependency parsing***
Supervisors:	**Prof. Sudeshna Sarkar**
Synopsis:	28 April 2020
Thesis Submission:	22 May 2020

2011-2013	**Master of Technology (M.Tech)**, *Computer Science and Engineering,* Indian Institute of Technology Kharagpur.
CGPA:	**8.60**
Thesis:	*Word Sense Disambiguation in Bengali language*
Supervisor:	Prof. Sudeshna Sarkar

2004-2008	**Bachelor of Technology (B.Tech)**, *Computer Science and Engineering,* National Institute of Technology Durgapur.
CGPA:	**8.29**
Thesis:	*Testing and Verification of ISCAS-85 benchmark circuits in C*
Supervisor:	Prof. Suchismita Roy

2003	**Indian School Certificate Examination (Class XII Board Exam)**.
Marks:	**82.25%**
Board:	Council for Indian School Certificate Examination, New Delhi

2001	**Indian Council for Secondary Education Examination (Class X Board Exam)**.
Marks:	**85.00%**
Board:	Council for Indian School Certificate Examination, New Delhi

Academic Appointments

Dec'2020 - Dec'2021	Birla Institute of Technology and Science Pilani (Hyderabad Campus)
Designation:	Assistant Professor
Dec'2021 - till date	Indian Institute of Technology (ISM), Dhanbad
Designation:	Assistant Professor

Publications

Journals:

1. **Ayan Das** and Sudeshna Sarkar; *"Transform, Combine, and Transfer: Delexicalized Transfer Parser for Low-resource Languages"*; in **ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)**, Vol. 19 Issue 1, No. 4, pp. 173-190, June 2019.

2. **Ayan Das** and Sudeshna Sarkar; *"A survey of the model transfer approaches to cross-lingual dependency parsing"*; in **ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)**, Vol. 19 Issue 5, No. 67, pp. 1-60, May 2020.

Conferences:

1. **Ayan Das** and Sudeshna Sarkar; *"A little perturbation makes a difference: Treebank augmentation by perturbation improves transfer parsing"*; in the proceedings of **16th International Conference on Natural Language Processing (ICON)**, Hyderabad, India, December 18–21, 2019.

2. **Ayan Das** and Sudeshna Sarkar; *"MorphBen: A neural morphological analyzer for Bengali language"*; in the proceedings of **20th International Conference on Computational Linguistics and Intelligent Text Processing (CiCLing)**, La Rochelle, France, April 7–13, 2019.

3. **Ayan Das**, Affan Zaffar, and Sudeshna Sarkar; *"Delexicalized transfer parsing for low-resource languages using transformed and combined treebanks"*; in the proceedings of **CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies**, pp. 182–190, Vancouver, Canada, August 3–4, 2017.

4. **Ayan Das**, Raghuveer Chanda, Smriti Agrawal and Sourangshu Bhattacharya; *"Distributed Weighted Parameter Averaging for SVM Training on Big Data"*; in the proceedings of **Workshop on Distributed Machine Learning at 31st AAAI Conference on Artificial Intelligence (AAAI)**, pp. 472–477, San Francisco, CA USA, February 4–5, 2017.

5. **Ayan Das**, Agnivo Saha and Sudeshna Sarkar; *"Cross-lingual transfer parsing from Hindi to Bengali using delexicalization and chunking"*; in the proceedings of **13th International Conference on Natural Language Processing (ICON)**, pp. 99–108, Varanasi, India, December 17–18, 2016.

6. **Ayan Das**, Pranay Yerra, Ken Kumar and Sudeshna Sarkar; *"A study of attention-based Neural Machine Translation models on Indian Languages"*; in the proceedings of **6th Workshop on South and Southeast Asian Natural language at the 26th International Conference on Computational Linguistics (COLING) 2016**, pp. 163–172, Osaka, Japan, December 11–12, 2016.

7. **Ayan Das**, Agnivo Saha and Sudeshna Sarkar; *"Development of a Bengali parser by cross-lingual transfer from Hindi"*; in the proceedings of **6th Workshop on South and Southeast Asian Natural language at the 26th International Conference on Computational Linguistics (COLING) 2016**, pp. 33–43, Osaka, Japan, December 11–12, 2016.

8. **Ayan Das** and Sudeshna Sarkar; *"Word Sense Disambiguation in Bengali applied to Bengali-Hindi Machine Translation"* in the proceedings of **10th International Conference on Natural Language Processing (ICON)**, Noida, India, 18-20 December, 2013.

9. Alapan Kuila, **Ayan Das** and Sudeshna Sarkar; *"A Graph Convolution Network-based System for Technical Domain Identification"* in the proceedings of **17th International Conference on Natural Language Processing (ICON)**, Noida, India, 18-21 December, 2020.

Research Projects

December 2016-	**Creation of a dependency treebank for Bengali language**.
Principal Investigator	Prof. Sudeshna Sarkar
Funding organization	Google Inc.
Role	I was involved in the development of the annotation guidelines for morphosyntactic features and dependency relations. I was involved in writing the project along with Prof. Sarkar. I have developed a PoS tagger, a lemmatizer, a morphological analyzer and dependency parser using the data developed for the project and setup a complete pipeline for automatic annotation of the data from raw text using the components stated above.

Tools developed

1. Developed an attention-based neural MT system from scratch using Theano library and studied the performance of this model on Indian language pairs.

2. Developed a joint neural part-of-speech tagger, morphological analyzer and lemmatizer for Bengali language.

Talks and Presentations

Talks

1. Invited talk at Workshop on Verbal Multiword Expressions in Indian Languages, ICON 2019, Hyderabad, India

Oral presentations

1. **Ayan Das**, Sudeshna Sarkar. "A little perturbation makes a difference: Treebank augmentation by perturbation improves transfer parsing", ICON 2019

2. **Ayan Das**, Agnivo Saha and Sudeshna Sarkar."Development of a Bengali parser by cross-lingual transfer from Hindi", WSSANLP (COLING) 2016

3. **Ayan Das**, Pranay Yerra, Ken Kumar and Sudeshna Sarkar."A study of attention-based Neural Machine Translation models on Indian Languages", WSSANLP (COLING) 2016

4. **Ayan Das**, Sudeshna Sarkar. "Word Sense Disambiguation in Bengali applied to Bengali-Hindi Machine Translation", ICON 2013

Poster presentation

1. **Ayan Das** and Sudeshna Sarkar. "Cross-lingual transfer parsing from Hindi to Bengali using delexicalization and chunking", ICON 2016

Teaching

1. Information Retrieval
2. Computer Vision
3. Machine Learning

4. C Programming Laboratory
5. Database Laboratory
6. OOP using JAVA Laboratory
7. Algorithm Design and Analysis Laboratory

Professional Experience

| Jul'2008 -
Jul'2011 | Tata Consultancy Services, Kolkata |
| Designation: | Systems Engineer |

Community involvement

Peer review of
journal paper

1. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)

Peer review
of conference
paper

1. International Conference on Natural Language Processing (ICON)

2. Empirical Methods in Natural Language Processing (EMNLP)

Awards

- Received "*Institute Travel Grant*" from IIT Kharagpur for attending COLING 2016.
- MHRD Scholarship from IIT Kharagpur for Doctoral Studies.

References

- **Prof. Sudeshna Sarkar**, Dept. of CSE, IIT Kharagpur,
  Email: *sudeshna@cse.iitkgp.ernet.in*

- **Prof. Suchismita Roy**, Dept. of CSE, NIT Durgapur,
  Email:*suchismita.roy@cse.nitdgp.ac.in*

- **Prof. Debdeep Mukhopadhyay**, Dept. of CSE, IIT Kharagpur,
  Email:*debdeep@cse.iitkgp.ac.in*

Declaration

I hereby declare that the above information is correct to the best of my knowledge