

Ashim Gupta

CONTACT INFORMATION

Artificial Intelligence Lab., CSE Dept. *Phone:* +91-8948561541
Indian Institute of Technology Kharagpur *Email:* ashimgupta95@gmail.com
West Bengal, India -721302 *Github:* ashim95

RESEARCH INTERESTS

Natural Language Processing, Computational Linguistics, Machine Learning, Deep Learning

EDUCATION

Indian Institute of Technology - BHU, Uttar Pradesh, India

Bachelor of Technology, Electrical Engineering

July 2012 - May 2016

- Undergraduate Coursework: Artificial Intelligence, Data Structures and Algorithms, Computer Systems, Programming in C, Microprocessor Architecture
- CGPA: **8.27**

RKG DAV Public School, Punjab, India

All India Senior School Certificate Examination (2012), Percentage: **94.2%**

All India Secondary School Examination (2010), CGPA: **9.6**

RESEARCH EXPERIENCES

Computer Science Department, IIT - Kharagpur

Project Officer, NLP Researcher

May, 2017 - present

Advisors : Prof. Sudeshna Sarkar, Prof. Pawan Goyal

- Working on Indian Government funded research project on Drug Repurposing through literature and patent mining
- Developed a fully contextualized biomedical NER system achieving state-of-the-art on 3 prominent datasets. (Submitted to ECIR, 2019)
- Machine Translation systems for morphologically rich low-resource Indian languages.

Computer Science Department, IIT - BHU, Varanasi

Undergraduate Researcher

Dec, 2014 - August, 2015

Advisors : Prof. Rajeev Srivastava

- Study of Content based image retrieval systems, relevance feedback, segmentation algorithms
- Work on mammogram retrieval, published at IEEE UPCON- 2015

ENGINEERING EXPERIENCE

Proptiger.com, Gurgaon, India

Software Developer

August, 2016 - March, 2017

- Development for Home Loan division of Proptiger. Gained experience working with Spring, JPA, MySql, Redshift. (Java)
- A real estate Price Estimator application, A generic crawling framework to scrap real estate data from multiple websites. (Python)

PUBLICATIONS

- **Gupta A.**, Goyal P., Sarkar S. *Fully Contextualized Biomedical Named Entity Recognition*. 41st European Conference on Information Retrieval(ECIR), 2019 [pdf]
- Pramanick M, **Gupta A.**, Mitra P. *An LSTM-CRF Based Approach to Token-Level Metaphor Detection*. FigLang Workshop at NAACL, 2018 [pdf]
- Singh VP., **Gupta A.**, Singh S., Srivastava R. *An Efficient Content Based Image Retrieval System for Normal and Abnormal Mammograms*. IEEE UPCON'15, IIT Allahabad [pdf]

SELECTED PROJECTS

- **Machine Translation for low-resource Indian languages** (Ongoing): An unsupervised Phrase-based machine translation system with initial phrase table induction using a bilingual

lexicon and iterative back-translation. Exploiting the use of an NMT initialized with synthetic data from PBSMT. BLEU Score of 7.0. (PyTorch, Moses MT)

- **Medical Scientific Text Classification using Hierarchical Neural Networks** (2018): A Hierarchical Bi-directional LSTM based system with attention for classification of PubMed abstracts along with a modification of CRF to incorporate sequence tagging information. Results superior to state-of-the-art method by 0.4 % on RCT 20k, 200k datasets (Tensorflow)
- **Multi-Sensor Data Fusion Using Kalman Filter** (2015-2016): A robust Kalman Filter algorithm to fuse data from a low-cost IMU with GPS in order to reduce the error in estimation of object position. (MATLAB)
- **Content based Mammogram Retrieval** (2015): Content Based Image Retrieval system for mass and calcification mammograms by developing a Modified region growing algorithm for segmentation of masses. (MATLAB)
- **Image Retrieval with Relevance Feedback** (2014): Retrieval of images from Corel database using Fuzzy c-means algorithm with relevance feedback mechanism using SVM. An Ensemble of clustering algorithms was implemented resulting in increased Precision and Recall. (MATLAB)

ACHIEVEMENTS AND AWARDS

- Cleared the highly prestigious **IIT-JEE (2012)**, and was placed among the top 0.5% from about half a million students
- Offered a SHE scholarship under Innovation in Science Pursuit for Inspired Research (INSPIRE).
- District and School topper in class 12th Board examination conducted by CBSE.
- Online Courses : Machine Learning by Stanford University, Introduction to Computer Vision by Georgia Tech., Introduction to Natural Language Processing by Stanford University.

VOLUNTEER TEACHING EXPERIENCE

Vindhya Gyan Public School, Mirzapur (near Varanasi) July, 2014 - May, 2016

- Taught and mentored students with non-English background from grade 6 to grade 12 on Saturdays, taught programming to students from grade 10 to grade 12.
- Organized various group and team-building activities on Sundays.

SKILL SET

- Programming Languages: Java; Python (Tensorflow, Keras, PyTorch, scikit-learn, NLTK); C++; MATLAB
- Technologies and Platforms: Linux, Git, MySql, Maven, Spring; Apache Solr, Redshift, L^AT_EX
- Good communication skills, collaboration and Interpersonal skills with proficiency in grasping new technical concepts quickly and utilizing them in an effective manner.