

# Nitish Gupta

---

PERSONAL INFORMATION 305 N. Lincoln Ave, Apt. 102  
Urbana, Illinois  
USA - 61820

e-mail: [ngupta19@illinois.edu](mailto:ngupta19@illinois.edu)  
<http://nitishgupta.github.io>  
Mobile: +1 - (847) 912 0112

EDUCATION **PhD Student in Computer Science** August 2015 – Present  
*University of Illinois, Urbana-Champaign*  
*Advisor : Dan Roth*

**B.Tech - M.Tech (Dual Degree) in Electrical Engineering** July 2010 – May 2015  
*Indian Institute of Technology Kanpur*  
*CGPA : M.Tech - 10.0/10.0, B.Tech - 8.8/10.0*

RESEARCH INTERESTS Machine Reading - *Knowledge Extraction from unstructured text, AKBC, Question Answering.*  
Large Scale Machine Learning

PUBLICATIONS **Revisiting the Evaluation for Cross Document Event Coreference**  
*Shyam Upadhyay, Nitish Gupta, Christos Christodoulopoulos, Dan Roth*  
*COLING, 2016*

**Collectively Embedding Multi-Relational Data for Predicting User Preferences**  
*Nitish Gupta, Sameer Singh*  
*arXiv preprint arXiv:1504.06165*  
**Grand Prize Winner of Yelp Dataset Challenge, Round 4**

INTERNSHIP **Understanding User Generated Text** May – August 2016  
*PhD Research Intern, Google Research, Mountain View, CA*

## AWARDS AND ACHIEVEMENTS

- Awarded **Grand Prize** in the **Yelp Dataset Challenge, Round 4**, 2015
- Awarded **Computer Science Excellence Fellowship, 2015**, UIUC
- **Department Rank 1** in the Masters Batch of 2015 among 120 students.
- Ranked in **Top 0.1%** (amongst 475,000 students) in IIT-JEE 2010
- Achieved **All India Rank 132** in National Science Olympiad 2010
- Awarded **CBSE Merit Scholarship** by Central Board of Secondary Education, India in 2010
- Received the **CBSE Merit Certificate**, 2008 for being among the **Top 0.1%** scorers in Science.
- Selected for **Visiting International Student Internship and Training(VISIT)** program, University of Washington in 2014

## RESEARCH EXPERIENCE

### “Unsupervised Neural Entity Linking”

*Advisor: Prof. Sameer Singh, UC Irvine. Prof. Dan Roth, UIUC (Ongoing)*

In this work, we propose a neural generative model for modeling natural language as being generated from a latent Knowledge Base. Our model is similar to a Discrete Variational Auto-encoder that we currently formulate for unsupervised entity linking. Being able to train on millions of unlabeled documents, we learn a strong natural language encoder that effectively represents contextual text for entity linking without the use of any labeled training data. Our model also allows to resolve entity mentions for entities that do not exist in the Knowledge Base thus implicitly modeling cross-document coreference (NIL Clustering, Cold-Start problem) that existing models do not facilitate directly. We also plan to extend this model for joint relation extraction. [*Work in progress*]

### “Learning Distributed Document Representations for Multi-Label Categorization”

*Masters Thesis - Advisor: Prof. Harish Karnick, CSE, IIT Kanpur*

This work focused on learning continuous distributed vector representations for documents such that documents with similar semantic content have similar vector representations in the low-dimensional space. We developed an unsupervised neural network model, based on the Skip-Gram model for learning embeddings. The model achieves state-of-the-art results on multi-label document categorization on the standard Reuters-21578 dataset. We also show how learned representations help in predicting missing categories. [[Thesis](#)]