

# Swarnadeep Saha

CS PhD Candidate, UNC Chapel Hill, NC, USA

🌐 <https://swarnahub.github.io>

🐙 <https://github.com/swarnaHub>

✉ [swarna@cs.unc.edu](mailto:swarna@cs.unc.edu) 📞 919-641-7845

## RESEARCH INTERESTS

---

Natural Language Processing, Machine Learning, Deep Learning, Reinforcement Learning.

## EDUCATION

---

**UNC Chapel Hill**

*Ph.D. in Computer Science, Advisor: Prof. Mohit Bansal*

**North Carolina, USA**

*2019 - Present*

**Indian Institute of Technology, Delhi**

*M.Tech. in Computer Science, GPA: 9.01/10.0*

**Delhi, India**

*2015 - 2017*

**Jadavpur University**

*B.E. in Computer Science, GPA: 8.72/10.0*

**Kolkata, India**

*2010 - 2014*

## EXPERIENCE

---

**UNC Chapel Hill**

*Research Assistant*

**North Carolina, USA**

*August 2019 - Present*

- Developing human-interpretable (graph-based, natural language explanation-based) models for downstream NLP tasks of Question Answering, Commonsense Reasoning, and Linguistic Formal Reasoning.

**IBM Research - India**

*Research Engineer*

**Bangalore, India**

*July 2017 - June 2019*

- Designing and implementing large scale **Machine Learning** and **NLP** solutions for **Intelligent Tutoring Systems (Watson Tutor)**, notably in the areas of **Automatic Short Answer Grading** and **Text Segmentation**.

**Adobe Systems India Pvt. Ltd.**

*Member of Technical Staff*

**Noida, India**

*June 2014 - July 2015*

- Worked as a full-stack software developer in the **Acrobat Reader Team** of Adobe India.

## PUBLICATIONS

---

**Archival Conference Papers**.....

1. **Swarnadeep Saha**, Sayan Ghosh, Shashank Srivastava, and Mohit Bansal “*PRover: Proof Generation for Interpretable Reasoning over Rules*”, In Submission at EMNLP 2020.
2. **Swarnadeep Saha**, Yixin Nie, and Mohit Bansal “*ConjNLI: Natural Language Inference over Conjunctive Sentences*”, In Submission at EMNLP 2020.
3. **Swarnadeep Saha**, Malolan Chetlur, Tejas I. Dhamecha, Shantanu Godbole and others “*Aligning Learning Objectives to Learning Resources: A Lexico-Semantic Spatial Approach*”, IJCAI 2019, Macau, China [Acceptance Rate: 17%].
4. Smit Marvaniya, **Swarnadeep Saha**, Tejas I. Dhamecha, Peter Foltz, Renuka Sindhgatta and Bikram Sengupta “*Creating Scoring Rubric from Representative Student Answers for Improved*

*Short Answer Grading*", CIKM 2018, Turin, Italy [Acceptance Rate: 17%].

5. **Swarnadeep Saha** and Mausam "Open Information Extraction from Conjunctive Sentences", COLING 2018, Santa Fe, New Mexico, USA [Acceptance Rate: 37%]
6. Tejas I. Dhamecha, Smit Marvaniya, **Swarnadeep Saha**, Renuka Sindhgatta and Bikram Sengupta "Balancing Human Efforts and Performance of Student Response Analyzer in Dialog-based Tutors", AIED 2018, London, UK [Acceptance Rate: 25%]
7. **Swarnadeep Saha**, Tejas I. Dhamecha, Smit Marvaniya, Renuka Sindhgatta and Bikram Sengupta "Sentence Level or Token Level Features for Automatic Short Answer Grading?: Use Both", AIED 2018, London, UK [Acceptance Rate: 25%]
8. **Swarnadeep Saha**, Harinder Pal and Mausam "Bootstrapping for Numerical Open IE", ACL 2017, Vancouver, Canada [Acceptance Rate: 18%]

#### Pre-prints and Workshops.....

1. **Swarnadeep Saha**, Tejas I. Dhamecha, Smit Marvaniya, Peter Foltz, Renuka Sindhgatta and Bikram Sengupta "Joint Multi-Domain Learning for Automatic Short Answer Grading", arXiv 1902.09183.

## MAJOR RESEARCH PROJECTS

---

### Proof Generation for Interpretable Rule Reasoning

UNC Chapel Hill

*Ph.D. Research [Supervised by Prof. Mohit Bansal]*

*Jan 2020 - Present*

- o Proposed PProver, an interpretable deep learning model that serves as a linguistic analog of formal reasoning by jointly answering questions and generating proofs.
- o PProver's proofs are highly accurate and also obtains state-of-the-art QA accuracy on the task.
- o Proof graphs are generated through a novel node and edge module of the model in the presence of multiple global constraints during training and ILP inference.
- o Currently working on developing a Deep Reinforcement Learning framework for the task that can generate human-like compositional proofs.

### Natural Language Inference over Conjunctive Sentences

UNC Chapel Hill

*Ph.D. Research [Supervised by Prof. Mohit Bansal]*

*August 2020 - December 2020*

- o Proposed the task and a dataset for Natural Language Inference over Conjunctive Sentences.
- o Conjunctions in English are challenging both syntactically and semantically, typically arising from their non-boolean usages.
- o Proposed new linguistically augmented models and training methods for tackling inferences over conjunctive sentences.

### Automatic Short Answer Grading for Intelligent Tutoring Systems

IBM Research - India

*Industry Research*

*July 2017 - June 2019*

- o Supervised machine learning models for ASAG require a lot of annotated data which is expensive and time consuming to collect. Proposed an iterative data collection and grading approach that balances human effort and performance of ASAG.
- o Traditional hand-crafted features and recent deep learning models have complementary benefits in ASAG. Developed a joint model that shows strong performance across datasets.
- o Human graders often follow a **Scoring Rubric** or **Mark Scheme** to grade student answers. Proposed a machine learning model that creates such a Scoring Rubric for improving the grading performance.
- o ASAG systems trained on one domain often suffer when tested on another domain. Developed an **end-to-end neural architecture** for **domain adaptation** of ASAG.

### Open Information Extraction from Numerical and Conjunctive Sentences

IIT Delhi

- State-of-the-art **Open Information Extraction (Open IE)** systems lose substantial recall due to ineffective processing of numerical and conjunctive sentences.
- Developed the first **Open Numerical Relation Extractor** using a bootstrapping technique.
- Developed a **Language Model based Coordination Analyzer** that splits conjunctive sentences into simple ones. Used this to improve open information extraction from conjunctive sentences.
- Released **Open IE 5.0**, the latest and widely used state-of-the-art software for Open Information Extraction.

## ACHIEVEMENTS AND AWARDS

---

- Awarded the **Munroe and Rebecca Cobey Fellowship** at UNC Chapel Hill.
- Awarded the **Best M.Tech Thesis** of 2015-2017 batch, CSE department, IIT Delhi.
- Awarded the **Research Appreciation Award** by IBM Research for work on Intelligent Tutors.
- Secured an **All India Rank of 142** in GATE, 2014.

## SOFTWARE SKILLS

---

- Programming Languages: C, C++, Java, Scala, Python, Perl, Assembly Languages.
- Databases: MySQL, PostgreSQL.
- Frameworks and Tools: Keras, PyTorch, Hadoop, Git, Perforce, Maven, SBT.

## RELEVANT GRADUATE LEVEL COURSES

---

- Machine Learning, Advanced Machine Learning, Graphical Models, Advanced NLP, Grounding in NLP, Machine Learning and Graphics.

## REFERENCES

---

- **Dr. Mohit Bansal**, Associate Professor, CS Department, UNC Chapel Hill.
- **Dr. Shashank Srivastava**, Assistant Professor, CS Department, UNC Chapel Hill.
- **Dr. Mausam**, Professor, CSE Department, IIT Delhi.