Swarnadeep Saha

CS PhD Candidate, UNC Chapel Hill, NC, USA

https://swarnahub.github.io
https://github.com/swarnaHub
swarna@cs.unc.edu 919-641-7845

RESEARCH INTERESTS

Natural Language Processing, Machine Learning, Deep Learning, Structured Prediction.

EDUCATION

UNC Chapel Hill — **North Carolina, USA**
Ph.D. in Computer Science, Advisor: Prof. Mohit Bansal — *2019 - Present*

Indian Institute of Technology, Delhi — **Delhi, India**
M.Tech. in Computer Science, GPA: 9.01/10.0 — *2015 - 2017*

Jadavpur University — **Kolkata, India**
B.E. in Computer Science, GPA: 8.72/10.0 — *2010 - 2014*

EXPERIENCE

UNC Chapel Hill — **North Carolina, USA**
Research Assistant — *August 2019 - Present*

- Developing human-interpretable models that can generate **structured explanations** for downstream NLP tasks of Question Answering, Commonsense Reasoning, and Linguistic Formal Reasoning.

IBM Research - India — **Bangalore, India**
Research Engineer — *July 2017 - June 2019*

- Designed and implemented 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. — **Noida, India**
Member of Technical Staff — *June 2014 - July 2015*

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

PUBLICATIONS

1. **Swarnadeep Saha**, Sayan Ghosh, Shashank Srivastava, and Mohit Bansal *"PRover: Proof Generation for Interpretable Reasoning over Rules"*, **EMNLP 2020** [Acceptance Rate: 24%].

2. **Swarnadeep Saha**, Yixin Nie, and Mohit Bansal *"ConjNLI: Natural Language Inference over Conjunctive Sentences"*, **EMNLP 2020** [Acceptance Rate: 24%].

3. Chul Sung, Tejas Dhamecha, **Swarnadeep Saha**, Tengfei Ma, Vinay Reddy, and Rishi Arora *"Pre-Training BERT on Domain Resources for Short Answer Grading"*, **EMNLP-IJCNLP 2019** [Acceptance Rate: 23%].

4. **Swarnadeep Saha**, Malolan Chetlur, Tejas I. Dhamecha, Shantanu Godbole and others *"Aligning Learning Objectives to Learning Resources: A Lexico-Semantic Spatial Approach"*, **IJCAI 2019** [Acceptance Rate: 17%].

5. 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** [Acceptance Rate: 17%].

6. **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**.

7. **Swarnadeep Saha** and Mausam *"Open Information Extraction from Conjunctive Sentences"*, **COLING 2018** [Acceptance Rate: 37%]

8. 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** [Acceptance Rate: 25%]

9. **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** [Acceptance Rate: 25%]

10. **Swarnadeep Saha**, Harinder Pal and Mausam *"Bootstrapping for Numerical Open IE"*, **ACL 2017** [Acceptance Rate: 18%]

MAJOR RESEARCH PROJECTS

Proof Generation for Interpretable Rule Reasoning **UNC Chapel Hill**
Ph.D. Research [Supervised by Prof. Mohit Bansal] *Jan 2020 - Present*

- Proposed PRover, an interpretable deep learning model that serves as a linguistic analog of formal reasoning by jointly answering questions and generating proofs.
- PRover's proofs are highly accurate and also obtains state-of-the-art QA accuracy on the task.
- 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.
- Currently working on developing a model that can generate multiple correct proofs.

Natural Language Inference over Conjunctive Sentences **UNC Chapel Hill**
Ph.D. Research [Supervised by Prof. Mohit Bansal] *August 2020 - December 2020*

- Proposed the task and a dataset for Natural Language Inference over Conjunctive Sentences.
- Conjunctions in English are challenging both syntactically and semantically, typically arising from their non-boolean usages.
- 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*

- 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.
- Traditional hand-crafted features and recent deep learning models have complementary benefits in ASAG. Developed a joint model that shows strong performance across datasets.
- 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.
- 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**

M.Tech Thesis [Supervised by Prof. Mausam] *April 2016 - December 2017*

- 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, Structured Prediction, 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.