Swarnadeep Saha

IBM Research - India, Bangalore, Karnataka - 560045

https://swarnahub.github.io
https://github.com/swarnaHub
writetoswarna@gmail.com +91-9643467463

RESEARCH INTERESTS

Natural Language Processing, Information Extraction, Textual Entailment, Machine Learning, Deep Learning, Intelligent Tutors.

EDUCATION

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

IBM Research - India — **Bangalore, India**
Research Engineer — *July 2017 - Present*

- 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. — **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

Archival Conference Papers

1. **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%].

2. 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%].

3. **Swarnadeep Saha** and Mausam *"Open Information Extraction from Conjunctive Sentences"*, COLING 2018, Santa Fe, New Mexico, USA [Acceptance Rate: 37%]

4. 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%]

5. **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%]

6. **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

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.

Automatic Short Answer Grading for Intelligent Tutoring Systems **IBM Research - India**
Industry Research *July 2017 - Present*

- 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.

TECHNICAL PROJECTS

- Developed a **Bot for the Game of Entropy** using **Minimax algorithm and Reinforcement Learning**.
- Developed a **Machine Learning** model to identify if two persons are same or not.
- Developed a **Multiplayer Soccer Game** using **OpenGL**.
- Developed an **ARM Simulator** supporting various complex instructions.

ACHIEVEMENTS AND AWARDS

- 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.
- Ranked **5th** in the M.Tech batch of 2015-2017, CSE department, IIT Delhi.
- 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 COURSES

- Artificial Intelligence, Machine Learning, Graphical Models, Advanced NLP, Data Mining.
- Databases, Software Systems Lab, Cloud Computing.
- Data Structures, Advanced Data Structures, Algorithms, Geometric Algorithms.

REFERENCES

- **Dr. Mausam**, Associate Professor, CSE Department, IIT Delhi.
- **Dr. Renuka Sindhgatta**, Lecturer, Queensland University of Technology.