Saurav **Saha**

Cognitive Science · Machine Learning · Python · Data Science · Artificial Intelligence

R-18, Boys Hostel 3, NIT Mizoram, Aizawl PIN : 796025, India

(+91) 9402508957 | contact.srvsaha@gmail.com | SRvSaha | srvsaha

"Be the change that you want to see in the world"

Education

National Institute of Technology Mizoram (NIT Mizoram) — *Aizawl, Mizoram, India*

B.Tech. in Computer Science and Engineering, Final Year — ***Aug. 2014 - June. 2018 (expected)***

- CGPA of **9.61/10** (till June 2017, 6th Semester)
- Consistenly maintaining **department rank 1** over past three years

S.R. Public Sr. Sec. School, Kota, Rajasthan (CBSE) — *Kota, Rajasthan, India*

High School — ***Apr. 2012 - May. 2014***

- Secured **93%** overall and ranked amongst top 5% in India
- 95% Maths, 95% Chemistry

Green Valley English School, Mankachar, Assam (SEBA) — *Mankachar, Assam, India*

Secondary School — ***Jan. 1999 - May. 2012***

- Secured **91.67%** in HSLC Examination, ranked **2nd in District** and **18th Rank Holder in State**
- 99% Science, 98% Advanced Maths, 94%- Maths

Conference/Workshop

2014 **SciPy India**, **International Conference** on Python for Scientific Computing organised by FOSSEE — *IIT Bombay*

2014 **NNSC**, **National Workshop** on Network Security & Implementation organised by Network Bulls — *IIT Bombay*

Experience

DAAD-WISE Scholar, Summer Research Intern, Universität Bremen, Germany — *Bremen, Germany*

Guide : Prof. Rainer Malaka, Head of Digital Media Group, TZI — ***Jun. 2017 - Aug. 2017***

- **Developed an Artificial Episodic Memory model for Robots for Mastering Everyday Activities.**
- **Field** : Natural Language Processing, Knowledge Representation, Artificial Intelligence, Semantic Web
- Researched on developing Artificial Episodic Memory using Ontology as Knowledge Base in Project EASE.
- Developed a Semantic Web as Knowledge base for Robot to store the episodes in memory
- Developed a Recursive Crawler from scratch to scrape kitchenstories.io website unofficially to prepare the dataset.
- Developed a pipeline that turns natural language instructions into individual concept sets corresponding to a given ontology.
- Developed an efficient discourse memory model from scratch which traces the occurrences of individual concepts over the entire set of instructions that encompass the specific everyday activity at hand.
- Researched on Narrative Enabled Episodic Memories(NEEMs) to enable cognitive capabilities, naive physics understanding and common sense reasoning for Robots
- **Link**: www.open-ease.org
- **Tools** : BeautifulSoup, LXML, requests, Protégé, OWL, Python, JSON, XML, Stanford CoreNLP, JAVA

Data Science Intern, IIT Mandi, Himachal Pradesh — *IIT Mandi*

Guide : Prof. (Dr.) Varun Dutt, IIT Mandi — ***Dec. 2016 - Jan. 2017***

- **Machine Learning and Data Mining project using Big Data in Health-Care**
- **Field** : Data Mining, Big Data, Machine Learning
- Developed a predictive model which could identify from EMR Datasets which patient is likely to buy which medicine using Machine Learning Techniques.
- Developed a predictive model for determination of Frequent/Infrequent buyer given the attributes of the patient.
- Case study and analyses of various databases like MongoDB, Cassandra, HBase, kdb+ to understand why kdb+ is better in handling realtime big data.
- Reduced the time required for training the system by using Weka-Parallel, a parallel computing architecture.
- Analyses of handling Big Data in Hadoop vs Spark for Machine Learning
- Built majority voted ensemble for binary-class and ternary-class classification task from scratch.
- Tuned the performance of Decision Tree ML Algorithm by hyperparameter optimization using GridSearch Algorithm.
- **Tools** : Python, Excel, Apache Hadoop, Weka, Weka-Parallel, kdb+

Paraphrase Detection in India Languages, FIRE-ISI 2016 *NIT Mizoram, Aizawl*

Guide : Dr. Partha Pakray, HoD CSE, NIT Mizoram, Mr. Sandeeep Dash, NIT Mizoram *Jul. 2016 - Sept. 2016*

- **Field** : Machine Learning, Textual Semantic Similarity
- Our system NLP-NITMZ is based on three features: Unigram Matching Ratio, Levenshtein Ratio and Cosine Similarity using Vector Space Model.
- Built two classifiers which can tag paraphrases, non-paraphrases and semi-phrases in Indian Languages, namely Hindi, Malayalam, Punjabi and Tamil. Our classifiers are voted ensambles built on the top of Naive Bayes, Support Vector Machines, Random Forest, Logistic Regression, J48 Machine learning algorithms and gives **95%+** accuracy in the Train Set. In Test Set, we got **91.55%** in Hindi, **83.44%** in Malayalam, **94.20%** in Punjabi and **83.44%** accuracy in Tamil.
- For Machine Learning portion we have used **Probabilistic neural network(PNN)** to predict the class.
- **Tools** : Python, JAVA, WEKA, MATLAB, XML
- **Publication** : Paper published on **8th meeting of Forum for Information Retrieval Evaluation (FIRE 2016), CEUR-WS.org/Vol-1737, Pages 256-259**

QA4FAQ - Question Answering for Frequently Asked Questions, EVALITA-2016 *NIT Mizoram, Aizawl*

Guide : Dr. Partha Pakray, NIT Mizoram, Head of NLP-NITMZ Research Team *Jul. 2016 - Sept. 2016*

- **Field** : Information Retrieval, Text Mining
- Since searching within the Frequently Asked Questions (FAQ) page of a web site is a critical task: customers might feel overloaded by many irrelevant questions and become frustrated due to the difficulty in finding the FAQ suitable for their problems.
- Developed a search-engine which can effectively retrieve a list of most relevant FAQs and corresponding answers related to the query issued by the user. Used Combinatorics approach for query and by rating each result fetched on a scale like 3,2,1, the most relevant one is shown.
- Our system can effectively give **97%** relevant search results based on the queries of the user which is much better than any prevalent IR methodologies.
- **Tools** : Python, JAVA, Nutch, Apache Tomcat, Italian Stop-word Corpus Building, Combinatorics, Page Rating & Ranking Algorithms
- **Publication** : Paper published on **3rd Italian Conference on Computational Linguistics (CLiC-it 2016)**

Winter Research Intern, Jadavpur University, Kolkata *Kolkata, India*

Guide : Prof. (Dr.) Dipankar Das, Jadavpur University *Dec. 2015 - Jan. 2016*

- **Phrase Extraction from English Sentences for Clausal Identification**
- **Field** : Information Extraction, Text Mining, Data Structures, Algorithms
- The system built on the top of Stanford Parser and NLTK can detect various type of Phrases and can separate them automatically which can be used to extract Clauses from texts.
- Developed a recursive algorithm based on stack data structure which keep track of the start and end of phrases within Phrases. The task is recursively solved to extract the phrases along with their type.
- **Tools** : Stanford Parser, NLTK, Python

Technical skills

Languages : Python(Primary), C, C++, JAVA, JSON, XML, Shell Script, SQL, PHP
Tools/Frameworks : Stanford Core NLP, Protégé, OWL, Semantic Networks, RDF, WEKA, NLTK, Stanford Parser, Nutch, Apache Tomcat, MySQL, kdb+ database, Oracle Database, git, vim, gdb, Sublime Text 3, Scikit-Learn, NumPy, Keras, Tensorflow, Sphinx, reStructuredText, Apache Hadoop, Microsoft Excel, Python unittest, LaTeX, Markdown, BeautifulSoup, PyQt, Scrapy
API : Twitter API
Platform : Linux(Primary), Windows, Mac

Additional Relevant Courses

Online/MOOC

2017	**Mining of Massive Datasets**, Prof. Jeff Ullman, Prof. Jure Leskovec, Prof. Anand Rajaraman	*Stanford University*
2017	**Introduction to Hadoop and MapReduce**, Sarah Sproehnle, Ian Wrigley, Gundega Dekena	*Cloudera*
2016	**Introduction to Machine Learning**, Prof. Andrew Ng	*Stanford University*

GIAN Courses

2017	**Deep Learning for Natural Language Processing**, Dr. Benoit Favre, Aix-Marseille University(AMU)	*France*
2016	**Natural Language Processing & Sentiment Analysis**, Prof. Alexander Gelbukh, Instituto Politécnico Nacional (IPN)	*Mexico*
2016	**Introduction to Robot Operating System**, Prof. David Pinto Avendaño, Benemérita Universidad Autónoma de Puebla (BUAP)	*Puebla, Mexico*

Achievements & Awards

Research

2016	**3rd Place**, Detecting Paraphrases in Indian Languages (DPIL), FIRE'16. Overall : 36 teams	*FIRE, ISI Kolkata*
2016	**5th Place**, Information Extraction from Microblogs Posted during Disasters, FIRE'16	*FIRE, ISI Kolkata*

Scholastic

2017	**Awarded DAAD-WISE Scholarship**, Research Internship in Universität Bremen, Germany	*DAAD, Germany*
2016	**Scored an absolute 10/10 grade** , Natural Language Processing & Sentiment Analysis	*GIAN, NIT Mizoram*
2016	**Scored an absolute 10/10 grade** , Introduction to Robot Operating System	*GIAN, NIT Mizoram*
2016	**Scored an absolute 10/10 SGPA**, Highest till date in CSE Dept, Spring Semester 2016	*NIT Mizoram*
2015	**Science Olympiad**, ALL MIZORAM RANK 4th in "MANTHAN", regional SCIENCE OLYMPIAD	*NIT Mizoram*
2014	**All India Topper**, Secured 99/100 in Physical Education in CBSE 2014, subject merit highest	*CBSE, India*
2014	**JEE MAIN**, Amongst the top **2.5%** in India, out of 1.4 million appeared candidates	*India*
2013	**All India Rank 6th**, NATIONAL SCIENCE CONCOURS organised by Maxscore in 2013	*Gurgaon, India*

Extra-Curricular

2016/17	**Gold Medallist**, Won the Gold Medal, best Badminton Player of NIT Mizoram in Shaurya 2016, 2017	*NIT Mizoram*
2015	**3rd Prize**, FASTEST RUBIK's CUBE solving competition in Anunaad'15	*NIT Mizoram*

Positions of Responsibility

2017	**Joint Secretary**, Organised various events and management of Morphosis, Annual Tech-Fest	*Morphosis, NITMZ*
2015/16	**Teaching Assistant**, Mentoring a group of 20 students giving hands-on exposure to C/C++ Programming and Competitive Programming	*NIT Mizoram*
2015	**Organiser**, Code Warrior, annual CODING contest of NIT Mizoram at ANUNAAD 2015.	*NIT Mizoram*
2014	**Senior Managerial Team Member**, Head of 31 NITs in "YUGMA", annual pan NIT magazine	*Inter NIT*
2014	**Founder**, Founder of HackerRank Club, NIT Mizoram (Coding Club of NIT Mizoram).	*NIT Mizoram*
2014	**Organiser**, Organised "WEB-o-TIC", the online marketing & publicity events in ANUNAAD'15.	*NIT Mizoram*
2014	**Campus Ambassador**, Campus Ambassador for Techniche, IIT Guwahati, Seismech, IIT Guwahati, Zigsaw Consultancy Services, HackerRank, PTBN - the Inter NIT Network	*Pan India*
2014	**Technical Co-ordinator**, Co-ordinated the Techno-Cultural Anunaad'15, responsible for Technical Events Management	*NIT Mizoram*

References

Dr. Partha Pakray
Head of the Department
Assistant Professor
International Co-operation Officer
Principal Investigator
Department of Computer Science & Engineering
National Institute of Technology (NIT) Mizoram
Aizawl - 796012 , India
parthapakray@gmail.com
www.parthapakray.com

Dr. Dipankar Das
Assistant Professor
Department of Computer Science & Engineering
Jadavpur University
Kolkata-700032
India
dipankar.dipnil2005@gmail.com
www.dasdipankar.com

Prof. (Dr.) Rainer Malaka
Research Professor & Head
Digital Media Lab
FB 03, TZI
Universität Bremen, Germany
Bremen - 28359, Germany
+49 421 218-64402
malaka@tzi.de
http://dm.tzi.de/staff/malaka/

Dr. Varun Dutt
Applied Cognitive Science Lab
School of Computing and Electrical Engineering
Assistant Professor
Indian Institute of Technology (IIT) Mandi
Kamand, Himachal Pradesh, India – 175 005
varun@iitmandi.ac.in
+91-1905-267041
http://faculty.iitmandi.ac.in/ varun/