

# Saurav Saha

NATURAL LANGUAGE PROCESSING · MACHINE LEARNING · PYTHONISTA · DATA SCIENCE · ARTIFICIAL INTELLIGENCE

44, Kacharipara, Mankachar, Dhubri, Assam, PIN : 783131, India

☎ (+91) 9402508957 | ✉ me@sauravsaha.in | 🌐 www.sauravsaha.in | 📱 SRvSaha

“Be the change that you want to see in the world”

## Research Interests

Artificial Intelligence, Machine Learning, Natural Language Processing, Artificial Neural Networks, Computer Vision, Information Retrieval, Sentiment Analysis, Sentic Computing, Image Processing, Human Computer Interaction, Semantic Similarity, Data Science, Cognitive Science, Deep Learning

## Education

### National Institute of Technology Mizoram (NIT Mizoram)

Aizawl, Mizoram, India

B.TECH. IN COMPUTER SCIENCE AND ENGINEERING, PRE-FINAL YEAR

Aug. 2014 - PRESENT

- CGPA of **9.55/10** (till June 2016, 4th Semester)
- Consistently maintaining **department rank 1** over past two 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

### 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, Data 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, JAVA

### 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 Extraction, 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. Perhaps they are right there, but just worded in a different way than they know.
- 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

## Paraphrase Detection in India Languages, FIRE-ISI 2016

NIT Mizoram, Aizawl

GUIDE : DR. PARTHA PAKRAY, HoD CSE, NIT MIZORAM, MR. SANDEEP DASH, NIT MIZORAM

Jul. 2016 - Sept. 2016

- **Field** : Machine Learning, Deep 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 built on the top of Naive Bayes, Support Vector Machine, Random Forest, Decision Tree, 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

## Information Extraction from Microblogs(Twitter) Posted during Disasters, FIRE-ISI 2016

NIT Mizoram, Aizawl

GUIDE : DR. PARTHA PAKRAY, NIT MIZORAM, HEAD OF NLP-NITMZ RESEARCH TEAM

Aug. 2016 - Sept. 2016

- **Field** : Social Media Code Mixing, Information Extraction, Information Retrieval
- Built a system that can deal with the noisy nature of microblogs(**TWITTER**) which are very small (at most 140 characters) and often written informally, using abbreviations, colloquial terms, etc, and effectively developed IR methodologies for extracting important information from microblogs posted during disasters.
- Our system can effectively show the most relevant results related to each topic(Tweets) with high precision Twitter data.
- **Tools** : Python, JAVA, Nutch, Apache Tomcat, Word2Vec, Page Ranking Algorithms, Twitter API, JSON, TF-IDF, Skip-gram, Continuous Bag of Words(CBOW)

## Technical skills

**Languages** : Python(Primary), C, C++, JAVA, JSON, XML, Shell Script, SQL

**Tools/Frameworks** : WEKA, NLTK, Stanford Parser, Nutch, Apache Tomcat, Oracle, Git, Vim, gdb, Sublime Text 3, MATLAB, Scikit-Learn, Numpy, Scipy, NetworkX(Python), Sphinx, reStructuredText,  $\text{\LaTeX}$ , Matplotlib, Scilab

**API** : Twitter API

**Platform** : Linux(Primary), Windows

## Relevant Courses

### ACADEMIC

2016	<b>Operating Systems</b> , Fall Semester	NIT Mizoram
2016	<b>Database Management Systems</b> , Fall Semester	NIT Mizoram
2015	<b>Introduction to Algorithms</b> , Spring Semester	NIT Mizoram
2015	<b>Data Structures</b> , Fall Semester	NIT Mizoram
2015	<b>Discrete Mathematics</b> , Fall Semester	NIT Mizoram
2015	<b>Graph Theory</b> , Fall Semester	NIT Mizoram
2015	<b>Numerical Methods, Statistics and Probability</b> , Fall Semester	NIT Mizoram

### ONLINE

2016	<b>Introduction to Machine Learning</b> , Prof. Andrew Ng	Stanford University
2016	<b>Artificial Intelligence</b> , Prof. Patrick Winston	MIT
2016	<b>Neural Networks for Machine Learning</b> , Prof. Geoffrey Hinton	University Toronto
2016	<b>Artificial Intelligence</b> , Prof. Deepak Kheemani	IIT Madras
2016	<b>Artificial Neural Networks and Applications</b> , Prof. S.Sengupta	IIT Kharagpur
2015	<b>Data Structures and Algorithms</b> , Dr. Naveen Garg	IIT Delhi
2015	<b>Natural Language Processing</b> , Prof. Dan Jurafski & Prof. Chris Manning	Stanford University
2015	<b>Natural Language Processing</b> , Prof. Pushpak Bhattacharyya	IIT Bombay

### GIAN COURSES

2016	<b>Introduction to Robot Operating System</b> , Prof. David Pinto Avendaño, Benemérita Universidad Autónoma de Puebla (BUAP)	Puebla, Mexico
2016	<b>Natural Language Processing &amp; Sentiment Analysis</b> , Prof. Alexander Gelbukh, Instituto Politécnico Nacional (IPN)	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

- 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 **Gold Medallist**, Won the Gold Medal, best Badminton Player of NIT Mizoram in Shaurya'16 *NIT Mizoram*  
2015 **3rd Prize**, FASTEST RUBIK'S CUBE solving competition in Anunaad'15 *NIT Mizoram*  
2015 **2nd Prize**, Counter Strike, Arcadia 2015 at ANUNAAD, the annual Techno-Cultural Festival *NIT Mizoram*

## Positions of Responsibility

---

- 2015 - **Teaching Assistant**, Mentoring a group of 20 students giving hands-on exposure to C/C++ *NIT Mizoram*  
Present **Programming and Competitive Programming**  
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  
☎ +91 8259065018  
✉ 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.) Alexander Gelbukh

Research Professor & Head  
Natural Language Lab, Centro de Investigación en Computación (CIC)  
Instituto Politécnico Nacional (IPN) Mexico  
Col. Nueva Industrial Vallejo  
Mexico City - 07738, Mexico  
☎ (+52 1) 55 1810-4587  
✉ gelbukh@gelbukh.com  
🏠 www.gelbukh.com