PRAVAL SHARMA

122D Avery Hall, Lincoln, NE 68588, USA | praval.s17@gmail.com | pravals17.github.io

EDUCATION

University of Nebraska-Lincoln, USA *August 2019 - August 2024 (Expected)*
Ph.D. Computer Science
Research topic: Event specific information extraction from structured text
Adviser: Dr. Ashok Samal and Dr. Leen-Kiat Soh

Mangalore University, India *August 2015 - July 2017*
M.Sc. Computer Science GPA: 9.03/10

Bangalore University, India *June 2011 - May 2014*
Bachelors of Computer Application GPA: 78.1/100

PUBLICATIONS

Sharma, P., Samal, A., Soh, LK., and Joshi, D. A spatially-aware data-driven approach to automatically geocoding non-gazetteer place names. ACM Trans. on Spatial Algorithms and Systems. 10 (1), (2024).

Sharma, P., Samal, A., Soh, LK., and Joshi, D. A spatially-aware algorithm for location extraction from structured documents. Geoinformatica 27, 645–679 (2023).

PROFESSIONAL EXPERIENCE

Graduate Research Assistant, University of Nebraska-Lincoln, USA *August 2019 - Present*
My research is primarily focused on developing novel algorithms for information extraction from text using spatial data mining and natural language processing.

- Developed novel algorithms to automatically extract event specific spatial information from structured documents and published papers in journals as primary contributor.
  - Developed a data-driven approach to automatically identify frequent patterns in the names of places in the form of affixes from a geographic context.
  - Analyzed place names from various countries (Germany, Afghanistan, Mozambique, Brazil, and Japan) to evaluate the generality of frequent patterns in place names across the world.
  - Demonstrated that places with similar affixes in their names are not randomly distributed in space using spatial statistical tools.
  - Developed a practical approach for continuously updating gazetteers for a geographic context by adding new place names at a much finer resolution.
- Developed an innovative approach to extract journalistic 5Ws (Where, When, What, Who, and Why) of the events described in structured documents using generative AI.
  - Designed a template based on the journalistic 5Ws of events.
  - Trained a large language model (LLM) to extract the journalistic 5Ws conditioned on the template and the documents.
- Developed a novel approach to identify the precise location of events described in structured documents using transformers and graph neural network (GNN).
  - Designed an approach to construct graphs using the place names mentioned in documents and their spatial relationships.
  - Developed a GNN-based algorithm that leverages the graph of place names to identify the precise event location.
- Designed an automated pipeline to scrape news reports and store them in a database using web scrapers and cron jobs to create a large corpora of news reports ($\approx$ 5 million) from seven different newspapers (four national and three regional) from India.
- Trained and led a team of nine human coders for text annotation and developed multiple ground truth datasets.
  - Created one of the largest manually annotated datasets of 1,000 documents for place name extraction that is statistically reliable, derived using multiple coders and verified using inter-coder

reliability measure.
  - Developed one of the largest manually annotated datasets of 10,000 documents for developing and evaluating event specific information extraction algorithms using multiple coders and verified its reliability using inter-coder reliability.

Graduate Teaching Assistant, University of Nebraska-Lincoln, USA *August 2023 - Present*
Courses: Computer Vision, Data Structures and Algorithms
- Assisted students with their queries during office hours.
- Graded homework and projects for 30-60 students and provided detailed feedback.

Software Engineer, Javra Software, Nepal *January 2018 - May 2019*
- Developed modules for projects **XEPST** (a project management software) and **LMS** (leave management system) using **PROGRESS 4GL**, **HTML**, **JavaScript**, and **CSS**.
- Worked on feature updates in the **X/Files** framework, an internal framework of the company on top of which software and web applications for clients are built.
- Worked on an extract, transform and load (ETL) tool development project using **Pentaho** and implemented OAuth 2.0 in the ETL tool for the company's client **Enza Zaden, Netherlands**.
- Implemented **Sticky Session** in the load balancer of the AWS server of Enza Zaden.

Software Engineer Intern, Leapfrog Technology Inc., Nepal *October 2017 - December 2017*
- Designed Web pages for the company using **HTML**, **JavaScript**, and **CSS**.
- Recreated popular small games (Ant Smasher, Flappy Bird) using JavaScript.

PROJECTS

Anticipating the Number of Social Unrest Event Using Spatial Models *March 2020-May 2020*
Developed a novel algorithm to predict the number of social unrest events in the state of Tamil Nadu in India utilizing predictor variables that denote the socio-demographic, economic, and climatology components of the study region using a spatial regression model. Analysis revealed that the climatology component (i.e., drought) did not contribute to social unrest in the study region. Used R.

Analysis of COVID-19 in New York State *March 2020-May 2020*
Analyzed the trend of spread of coronavirus based on factors such as population density and social vulnerability at county level in the state of New York from 03/01/2020 to 04/14/2020 and identified the hotspot regions in the state using Getis-Ord Gi* statistic. Computed an index to denote the risk of getting infected by the virus in each county of the state using the social vulnerability and new case rate variables. Used ArcGIS Pro.

Deep Learning for Object Counting in Agriculture *October 2019-December 2019*
Developed a novel deep learning-based plant phenotyping model using convolution neural network (CNN) to count the number of leaves in maize and sorghum plants. The algorithm had three CNN layers and two dense layers. Following each CNN layer, there was a sequence of a dropout layer, a batch normalization layer, and a max pool layer with a stride of two. Used Python and TensorFlow.

Predictive Models for House Price Estimation: Data Mining Approach *January 2017-July 2017*
Evaluated the performance of machine learning models, including lasso regression, ridge regression, elastic net regression, and XGBoost, in estimating house prices using the Ames Housing dataset. Results indicated that lasso regression outperformed the other predictive models. Used Python.

RESEARCH PRESENTATIONS

Seminar talk *July 2020*
SNR Summer Seminar Series, University of Nebraska-Lincoln
Topic: Anticipating the Number of Social Unrest Event Using Spatial Models.

Research demo and presentation *November 2019*
Graduate Information Day, University of Nebraska-Lincoln
Topic: Approaches for Extracting Journalistic 5Ws of Events Reported in Structured Documents.

OUTREACH ACTIVITIES

Student Panelist: Graduate Information Day, University of Nebraska-Lincoln *Spring 2023*
Discussed my experience as a graduate student and answered questions from prospective students.

Student Project Stakeholder: Senior Design Project, University of Nebraska-Lincoln *Fall 2021*
Drafted a project description document for development of a web application for map-based visualization of spatio-temporal information about events reported in news reports and acted as the main point of contact for a team of six students during the development.

MENTORING

Undergraduates Ritvik Handa and Savan Patel, University of Nebraska-Lincoln *Fall 2020*
Undergraduates Pranav Nikam and Akshita Goel, University of Nebraska-Lincoln *Fall 2021, Spring 2022*

TECHNICAL SKILLS

Python, PyTorch, Tensorflow, PyG, Hugging Face, R, ArcGIS Pro, QGIS, HTML, CSS, JavaScript, PROGRESS 4GL, MySQL, Bash, Git

AWARDS AND SCHOLARSHIPS

Graduate Research Assistantship, University of Nebraska-Lincoln, USA *2019 - Present*
Graduate Teaching Assistantship, University of Nebraska-Lincoln, USA *2023 - Present*
Nepal Bidhya Bhusan, Government of Nepal *2018*
Award conferred to Nepali citizens with outstanding performance in their graduate degree.
First Rank, Mangalore University *2018*
Award bestowed to students securing the highest grade in their department in the university.
Silver Jubilee Scholarship, Government of India *2015*
Scholarship covered the full cost of study for a graduate degree in India.

MEMBERSHIP

ACM

REFERENCES

Ashok Samal
Professor of Computer Science
University of Nebraska-Lincoln

Leen-Kiat Soh
Professor of Computer Science
University of Nebraska-Lincoln

Deepti Joshi
Associate Professor of Computer Science
The Citadel, South Carolina