

# Debjyoti Paul

---

CONTACT 130 S 800 E APT 101  
Salt Lake City  
ZIP: 84102  
University of Utah  
USA

🏠 [www.cs.utah.edu/~deb](http://www.cs.utah.edu/~deb)  
✉ [deb@cs.utah.edu](mailto:deb@cs.utah.edu)  
in [/in/debjyotipaul385/](https://in.linkedin.com/in/debjyotipaul385/)  
🔗 [www.github.com/debjyoti385](https://www.github.com/debjyoti385)  
☎ +1 385 E1E-7219

RESEARCH INTERESTS Spatio-Temporal Data Analysis, Social Media Analysis, Healthcare Analytics, Machine Learning, Representation Learning, Data Visualization

EDUCATION [2015-2019] **PhD., University of Utah, School of Computing**, Salt Lake City, UT, USA  
Major : Computer Science & Engineering  
Advisor : *Prof. Feifei Li*  
Thesis : *Spatio-temporal Data Analysis, Prediction & its Applications.*  
GPA : 3.94/4.0  
[2011-2013] **Masters of Technology, Indian Institute of Technology Kanpur**, Kanpur, INDIA  
Major : *Computer Science & Engineering*  
Advisor : *Late Professor Sanjeev K Aggarwal*  
Thesis : *Multi-constraint Job scheduling in Grid Computing*  
GPA : 8.67/10.0 (Rank: 3)  
[2007-2011] **Bachelors of Technology, West Bengal University of Technology**, Kolkata, INDIA  
Major : *Computer Science & Engineering*  
College : *Institute of Engineering and Management*  
CGPA : 8.93/10 (Rank: < 10)

WORK EXPERIENCE [2018] **Facebook, Inc.**, Research Summer Internship, Seattle, United States  
- Apply various Machine Learning techniques for Search Ranking improvements.  
[2017] **Amazon AI Research**, Research Summer Internship, New York, United States  
- Hyperparameter optimization with MxNet.  
[2013-2015] **Flipkart**, Data Engineer, Bangalore, INDIA  
- Served as data engineer at data platform team *flipkart.com*, India's biggest e-commerce company.  
- Facilitated scalable environment for Big data analytics with processing pipelines for batch and stream.

THESES [2013] **M.Tech Thesis, Advisor: Late Dr. Sanjeev Kumar Aggarwal**  
**Multi-constraint Job scheduling in Grid Computing**  
The objective is to efficiently schedule Jobs on Grid to achieve maximum utilization of resources with energy efficient approach. Modeling real world computing & storage grid on Multi-objective Evolutionary Algorithm satisfying hard constraints on jobs constraints, resources constraints, and soft constraints on cost, and energy consumption on NSGA-II. Local optimization using Pareto optimal front, global optimization by applying mutation on population on search space. Also introduced job grouping technique with constraints to accumulate fine grained jobs which keeps processing time as low as possible.

PUBLICATIONS  
+ Bursty Event Detection Throughout Histories, **Debjyoti Paul**, Yanqing Peng, Feifei Li, *35th IEEE International Conference on Data Engineering (ICDE 2019)*, Macau, China, 2019.  
+ Geotagged US Tweets as Predictors of County-Level Health Outcomes, 2015–2016, Quynh C. Nguyen, Matt McCullough, Hsien-wen Meng, **Debjyoti Paul**, Dapeng Li, *American Journal of Public Health, September, 2017, DOI: 10.2105/AJPH.2017.303993*.  
+ Compass: Spatio Temporal Sentiment Analysis of US Election, **Debjyoti Paul**, Feifei Li, Murali Krishna Teja, Yu Xin, Richie Frost, *What twitter says!, 23rd SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD 2017)*, Aug 13-17, 2017, Halifax, Canada. DOI: 10.1145/3097983.3098053.  
+ Social media indicators of the food environment and state health outcomes, Quynh C. Nguyen, Hsien-wen Meng, Dapeng. Li, Matt McCullough, **Debjyoti Paul**, Kanokvimankul. P, Nguyen. T, Li. Feifei, *American Public Health Association, 148, 120-128.*, 2017, DOI: 10.1016/j.puhe.2017.03.013.

- + Multi-objective Evolution based Dynamic Job Scheduler in Grid, **Debjyoti Paul**, Sanjeev K. Aggarwal, *The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2014)*, IEEE, July 2nd - 4th, 2014, Birmingham, UK. DOI: 10.1109/CISIS.2014.50.
- + RCached-tree: An Index Structure for Efficiently Answering Popular Queries, Manash Pal, Arnab Bhattacharya, **Debjyoti Paul**, *ACM International Conference on Information and Knowledge Management (CIKM 2013)*, Oct. 27 - Nov. 1, 2013, San Francisco, CA, USA. DOI: 10.1145/2505515.2507817.
- + Lightweight Security Enhancement Protocol for Radio Frequency Identification (RFID), *IEEE SPSITM International Conference*, **Debjyoti Paul**, Sumana Basu, Punit Beriwal, IEEE, 2011, Kolkata, WB, INDIA

PROJECTS  
(FOR MORE...  
🏠/#PROJECTS)

Compass  
[estorm.org](http://estorm.org)

This project explores how Twitter can be used to analyze the popularity of Political Parties for the Presidential Election 2016. It is purely based on the data collected from Twitter. Read our Compass paper for details.

Event Aggregator

This project finds events from news articles, categorize them and gathers all articles talking about same event to a set. Each set of articles is referred as Event Entity. Information extraction process is applied to extract more information related to it. The code can process large scale data. The code will be made public later.

QuakeAnalysis  
[goo.gl/ovcWVT](http://goo.gl/ovcWVT)

This project is based on the seismic activity across world which widely varies in characteristic and patterns. We have found some distinguish patterns among the seismic activities and present them in an insightful manner. Check [demo](#) and [wiki](#) of Exploration & Analysis

MusicAtlas  
[goo.gl/UH54UP](http://goo.gl/UH54UP)

This website is designed for music lovers to learn about music based on countries. This is a unique tool to explore and analyze the trend of music based on time frame, genre and artists. Almost all data from 19th century to till date.

QuestionAnswering  
[goo.gl/QGfW2y](http://goo.gl/QGfW2y)

A Natural Language Processing project focussed on closed domain Question-Answering System. The project has Question Classifier, Sentence Similarity, Answer formulation and Coref resolution modules. The system has *Recall* of 63% and *F-score* of 43%.

Dartnews  
[goo.gl/wMOL4C](http://goo.gl/wMOL4C)

A street news browsing application, with an interactive GIS interface. News is organized by locality and topic. The user can explore based on topics (eg crime, politics etc) and geolocation. We used Context Dependent Geoparsing, where we attempt to find out which location is relevant to the News. Across all the topics, we saw at least 87% accuracy of topic prediction, and at least 80% accuracy of location prediction.

RCached Tree  
[doi.org-rcached-tree](http://doi.org-rcached-tree)

The objective is to speed up performance of point, range and kNN search queries for popular queries in databases. We came up with a variant of the R-Tree indexing structure with some features like caching at each node for popular queries.

Intelliad  
[goo.gl/UAnEdz](http://goo.gl/UAnEdz)

A Social Media driven Intelligent Ad-Targeting framework using Geo-profiling. The idea is to tag all geo-location enabled tweets(available publicly) with semantic categories (say sports, politics etc.) and their sentiment (positive, neutral, negative) using text mining. To enable serving of Ads, they also need to be tagged using same categories based on their content.

HACKATHON  
TROPHIES

[2017] HackTheU	HackTheU MLH Hackathon Winner (around 40 teams)
[2016] EMC2 Code	Mars Challenge Hackathon Winner (around 45 teams)
[2015] Goldman Sachs	Air Quality Hackathon 1st Runners-up (around 30 teams)
[2014] InMobi	Freedom Hack Worldwide Hackathon 1st Runners-up (around 160 teams)
[2013] Yahoo	Yahoo HackU 2013 Hackathon Winner (around 40 teams)

ACHIEVEMENTS

- [2012] Secured All India Rank 228 in GATE 2012 among 0.16 million participants and 223 in 2011 among 0.15 million participants of Computer Science & Information Technology department.
- [2011] Achieved All India Rank of 7 in Indian Space Research Organization (ISRO) recruitment exam for Junior Research Fellow among 0.2 million candidates.