Debjyoti Paul

CONTACT INFORMATION

130 S 800 E
Salt Lake City
Pin: 84102
University of Utah
USA

+1 385 313-7219
deb@cs.utah.edu
debjyoti.paul@utah.edu
http://github.com/debjyoti385
http://www.cs.utah.edu/~deb

RESEARCH INTERESTS

Large-scale data analytics using machine learning and data mining techniques, social media data analytics, mining and learning for social science data.

EDUCATION

University of Utah, School of Computing, Salt Lake City, UT, USA

PhD. Student, Computer Science & Engineering, Fall 2015- *Current*

Current Research Advisor: Professor Feifei Li
GPA: 3.95/4.0

Indian Institute of Technology Kanpur, Kanpur, UP, INDIA

M.Tech., Computer Science & Engineering, 2011-2013

Thesis Title: *Multi-constraint Job scheduling problem in Grid*
Research Adviser: Professor Sanjeev K Aggarwal
Research: Efficient scheduling strategy of Jobs on Grid with energy efficient solution for defined SLAs.
CGPA: 8.67/10 (*Rank: 3*)

West Bengal University of Technology, Salt Lake, Kolkata, WB, INDIA

B.Tech., Computer Science & Engineering, 2007-2011

College: *Institute of Enginering and Management*
CGPA: 8.93/10 (*Rank:* < 10)

WORK EXPERIENCE

Flipkart, July 2013 - May 2015, Bangalore, INDIA

Software Developer in Data Platform (Exceeds Expectations for performance in Jan-July 2014)

- Data Platform is data bank of *flipkart.com*, India's biggest e-commerce company
- Experience in building Data Processing Platform (Ingestion-Processing-Visualization) and data warehousing.
- Practied Big data analytics in scalable environment, processing pipeline for batch and stream. Database administrator of HP Vertica.

THESES

M.Tech Thesis, Advisor: Late Dr. Sanjeev Kumar Aggarwal

Multi-constraint Job scheduling problem in Grid

- The objective is to efficiently schedule Jobs on Grid to achieve maximum utilization of resources with energy efficient approach. Modelling 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.
- Finding near Optimal Scheduling Solution: Local optimization using Pareto optimal front, global optimization by applying mutation on population on search space. Introduced job grouping technique with constraints to accumulate fine grained jobs which keeps processing time as low as possible.

PUBLICATIONS

Debjyoti Paul, Sanjeev K. Aggarwal, Multi-objective Evolution based Dynamic Job Scheduler in Grid, *The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2014)*, IEEE, July 2nd - 4th, 2014, Birmingham, UK.

Manash Pal, Arnab Bhattacharya, **Debjyoti Paul**, RCached-tree: An Index Structure for Efficiently Answering Popular Queries, *ACM International Conference on Information and Knowledge Management (CIKM 2013)*, Oct. 27 - Nov. 1, 2013, San Francisco, CA, USA.

Debjyoti Paul, Sumana Basu, Punit Beriwal, IEEE Paper titled Multilevel Security Protocol using Radio Frequency Identification (RFID), *International Conference on Emerging Trends in Mathematics and Computer Applications2010*, *pg*-544-547, 2010, Sivakasi, TN, INDIA.

Debjyoti Paul, Sumana Basu, Punit Beriwal, IEEE, *Lightweight Security Enhancement Protocol for Radio Frequency Identification (RFID) in SPSITM International Conference*, 2011, Kolkata, WB, INDIA

Projects (for more..)

Twitter Election 2016 Sentiment Analysis -*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. We provide unbiased analysis of popularity based on the sentiment analysis for each party (Republican and Democratic) at basic geospatial unit area which is 'county' (not Electoral Vote region) is our case. We also present time-range analysis of the data. Also a Bursty Event Detection System is used to detect when surge of tweets happened. Our Bursty Event detection correctly identified the tweet bursts and associated with the event happened at that time.

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 referredas 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 after the paper is published.

QuakeAnalysis -*https://goo.gl/tp3sEk*
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 -*https://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. Please make sure you allow SSL connection from api server via this link before accessing @musicatlas.

QuestionAnswering -*https://goo.gl/QGfW2y*
Closed domain Question Answer System. The system has *Recall* of 63% and *F-score* of 43%.

AirQuality -*https://goo.gl/F8twhn*
Goldman Sachs Hackathon - This analyze the air quality of Utah state from real world dataset. Data warehousing and cubing to enable interactive visualization framework, which helped in determining pattern of Pollution and its reason.

RCachedTree -*10.1145/2505515.2507817*
The objective is to speed up performance of point, range and kNN search queries for frequent or 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 -*https://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.

Dartnews -*https://goo.gl/f8imWH*
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.

Hackathons

[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)

Skills

Programming — Python, Java, C, C++, Javascipt, D3.js, Three.js

Databases *etc.* — MySQL, HP Vertica, HiveQL, VSQL, Apache Pig, HBase, MR2, Apache Storm

Tidbits — LaTeX, Shell script, gdb, CSS3, HTML5, Maven, Vim, Azkaban2 *etc.*

Achievements

[2013] Ranked 3rd out of 39 M.Tech students of CSE department in Indian of Institute Technology, Kanpur

[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.