Debjyoti Paul

CONTACT INFORMATION

130S 800E
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

Databases, Big Data, Data mining, Distributed Systems, Indexing and searching in Databases

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
CPI: 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 to achieve maximum utilization of resources with energy efficient solution and maintaining hard and soft constraints along with 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 Development Engineer in Data Platform

- Data Platform is data bank of *flipkart.com*, India's biggest e-commerce company
- Experience in building Data Processing Platform (Ingestion-Processing-Visualization).
- Proficient in data warehousing concepts
- Practied Big data analytics in scalable environment, processing pipeline for batch and stream. Database administrator of HP Vertica.
- Rated as Exceeds Expectations for performance in Jan-July 2014.

SUMMER SCHOOLS

[2012] Microsoft Research, *Summer School on Distributed Computing*
Bangalore, India
- Familiarization with current research work on distributed computing

[2010] Jadavpur University, *System Administration*, Trainee,
Kolkata, India
- Practical experience with servers and networking devices and on various aspect of system administration

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

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/ formularizing 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.

Scholarship

- Research Fellowship from School of Computing, University of Utah worth around $50,000 for 1st year of Ph.D.
- GBP 1,700 from IIT Kanpur for Paper presentation in CISIS 2014, Birmingham, UK
- 96,000 INR per year from Ministry of Human Resource Development India during M.Tech.
- 2 years waiver of tuition fees at school for academic excellence in school.

Skills

Programming	Python, Java, C, C++, Java
Scripts and Query languages	LaTeX, Shell script, Javascript, Pig, VSQL, HiveQL (Custom UDFs), SQL
Databases	MySQL, HP Vertica, HBase
Tech tools and Softwares	MapReduce2, Apache Storm, gdb, Vim, OhMyZsh, IntellijIDEA, Maven
Tidbits	CSS3, HTML5, Data Visualization (d3.js, Three.js), Azkaban2 & Oozie exec engine

Projects

Twitter Election 2016 Sentiment Analysis

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.

MusicAtlas

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.

QuakeAnalysis

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 wiki of Exploration and Analysis

QuestionAnswering

Closed domain Question Answer System. The system has *Recall* of 63% and *F-score* of 43%.

AirQuality @Utah

Goldman Sachs Hackathon Runners-Up. Explored and analyzed air quality of Utah from given dataset. Applied data warehousing strategies and built a cubing server to serve the front end. Frontend includes interactive visualization framework to analyze Air Quality / Pollution in Utah region.

Metonym (metonym.in)

A synonym based vocabulary learner website built with d3.js, php and dropwizard framework at the backend. For demo use username: "demo" & password: "12345678"

Intelliad

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. In the end, for every geolocation-enabled ad request, tags of data points in vicinity (the neighbourhood being controlled online) will be used to generate a runtime profile for the request, and appropriate ads will be picked for serving.

Projects (continued)

Dartnews — A street news browsing application, with an interactive GIS interface. News is organized by locality and topic. The user can choose one or more topics of interest (eg crime, politics etc), and also narrow down to the locality of interest. This was done using a technique that we call 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 — 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.

Elliptic Curve Diffie-Hellman Key Agreement Protocol — We here proposed an efficient Elliptic curve Diffie-Hellman two-party key agreement protocol using public key authentication based on Elliptic Curve Group. We presented an improved version of that protocol which uses Converse of Fermats theorem with eliminated charmichael numbers. Also efficiently generating primitive roots for D-H key.

Real time discrimination of Speech and Music — Real-time speaker segmentation is required in many applications, such as speaker tracking in real- time news-video segmentation and classification, or real-time speaker adapted speech. After feature extraction, using Low Short Term Energy Ratio (LSTER) the input digital audio stream is classified into speech and music. Here we have used DSK6713 Digital Signal Processing Kit.

Scholastic Achievements

- **[2016]** EMC2 Code Mars Challenge Hackathon Winner 2016 (out of 50 teams) event organized at University of Utah, Salt Lake City, UT (Team of 4)
- **[2015]** 1st Runner-up at Goldman Sachs Utah Hackathon event organized in University of Utah, Salt Lake City, UT (Team of 3)
- **[2015]** Acknowledged Best Poster Presentation for Natural Language Processing course CS6340 at University of Utah.
- **[2014]** 1st Runner-up among 160 teams in Freedom Hack worldwide hackathon event organized by InMobi India, Bangalore in memory of Aaron Swartz (Team of 4)
- **[2013]** Ranked 3rd out of 39 M.Tech students of CSE department in Indian of Institute Technology, Kanpur, M.Tech
- **[2012]** Awarded the 1st prize in the application development competition HACKU 2012, organized by Yahoo! At IIT Kanpur.
- **[2012]** Secured All India Rank 228 in GATE 2012 among 0.16 million participants of Computer Science & Information Technology department.
- **[2012]** Acknowledged as the Best Project by the course professor for Boosting performance of popular queries which was done as part of CS618 (Indexing and Searching of Databases) course.
- **[2011]** Achieved All India Rank of 7 in Indian Space Research Organization recruitment exam for Junior Research Fellow among 0.2 million candidates.
- **[2011]** Secured All India Rank 223 in GATE 2011 among 0.15 million participants of Computer Science & Information Technology department.
- **[2007-2011]** Amongst top 10 student of CSE department in Institute of Engineering Management, Kolkata, and awarded academic excellence for performance in B.Tech.
- **[2010]** 2nd Rank in Manual Robo-race competition & 2nd Rank in Robo-Olympic an innovative track based racing competition in Bits to Bytes Tech-fest, Bengal Institute of Technology.
- **[2007]** Secured 99.2 percentile in State level Engineering Entrance Exam (WBJEE)

Position of Responsibilities

[2016] Teaching Mentor for course CS5340-Natural Language Processing at University of Utah for Fall 2016.

[2015-*current*] Member of Data group University of Utah.

[2012] Member and organizer of SIGDATA, a Special Interests Group on DATA which meets weekly to discuss techniques related to data management and mining at IIT Kanpur.

[2012-2013] Teaching assistant for Software Engineering (CS455) course at IIT Kanpur.

[2011-2012] Teaching assistant for Fundamentals of Computer Science (ESC101) course at IIT Kanpur.

[2009] Event coordinator and Website developer for Festronix Techfest at Institute of Engineering and Management

References

Available upon request