

Arindam Paul

CONTACT INFORMATION

2145 Sheridan Road
Department of EECS
Northwestern University
Evanston, IL 60208 USA

Phone: (440) 622-1087
E-mail: arindam.paul@eecs.northwestern.edu
Website: www.arindampaul.me

INTERESTS

Machine Learning, Materials Informatics, Information Retrieval, Data Mining, Natural Language Processing, Social Media & Networks, Human Computer Interaction

EDUCATION

Northwestern University, Evanston, Illinois

Ph.D. Candidate, Computer Engineering (Expected graduation date: June 2018.) [GPA: 3.6 / 4.0]
Advisors: Prof. Alok Choudhary, Prof. Ankit Agrawal

Northwestern University, Evanston, Illinois

Master of Science, Computer Science, Summer 2014 [GPA: 3.6 / 4.0]

- Course Highlights: Machine Learning, Natural Language Processing, Social Network Analytics, Artificial Intelligence, Algorithms, Online Advertising Systems, Advanced Networks, Distributed Systems, Quantitative Research Methods, Operating Systems.

Birla Institute of Technology & Science, Pilani, Rajasthan India

Master of Engineering (Hons.), Software Systems, May 2012 [GPA: 7.9 / 10.00]

- Dissertation: Designing an efficient Distributed Computing Solution for Data Mining
- Course Highlights: Data Mining, Data Warehousing, Database Management Systems, Software Engineering & Management, Software Architecture, Pervasive Computing, Object Oriented Analysis & Design, Data Storage Technologies & Management, Computer Organization, Data Structures, Object Oriented Programming (Java)

Birla Institute of Technology & Science, Pilani, Rajasthan India

Bachelor of Engineering (Hons.), Chemical Engineering, Dec 2009 [GPA: 7.55 / 10.00]

- Thesis: Detecting Sybil Attacks in P2P networks using Psychometric Analysis
- Course Highlights: Structure and Property of Materials, Mechanics of Solids, Process Engineering and Design, Optimization, Operations Research, Numerical Analysis, Control Systems, Chemical Kinetics

ACADEMIC EXPERIENCE

Northwestern University, Evanston, Illinois USA

Graduate Researcher

Fall 2012 - present

Is/Was involved in the following research projects.

Facebook Confessions & Yik Yak

- Studied question asking about sensitive topics in anonymous forums
- Designed a system to automatically identify and classify taboo posts in anonymous forums with good accuracy

Learning from Ads: Reverse-engineering demographics and interests

- Created synthetic user profiles with different demographic and interest features and collecting ad traffic
- Created a model by using cross-validation which can predict user features from resulting data-set and ground-truth
- Application of the model to cellular web-data to predict user's demographics and interests on-the-fly

Teaching Assistant

Winter & Spring 2014, Winter & Fall 2015

Assisted the instructor in teaching the following undergraduate level courses. Duties included sharing of responsibilities for lectures, exams, homework assignments, grades, office hours and leading computer lab exercises.

- EECS 110 Introduction to Computer Programming (Python) , Winter & Spring 2014
- EECS 110 Introduction to Computer Programming (Python) , Winter 2015
- EECS 214 Data Structures and Data Management, Fall 2015

Guest Lecturer
EECS 510 Social Media Mining

Spring 2016

- *Impact of 'likes' and 'reactions' on social media*
- *Anonymity in Social Media*
- *Crawling and scraping the web*

Instructor

Summer 2016

MGLC Transferable Skills Workshop on Machine Learning

Attended by McCormick Graduate students and faculty

- *What and Why of Machine Learning ?*
- *Algorithms*
- *Application*
- *Introduction to Deep Learning*

BITS Pilani, Rajasthan, India

Graduate Researcher

Spring 2010 - Spring 2012

Was involved in the following research projects.

Designing an efficient Distributed Computing Solution for Data Mining

Fall 2011 -Spring 2012

- Created a Beowulf Linux(Ubuntu) cluster using OpenMPI library project
- Implemented parallel implementation of K-means for OpenMPI
- Bench-marked sequential and parallel OpenMPI implementations of K-means clustering algorithm
- Compared the performance with the control Hadoop cluster

Preventing Sybil attacks in P2P systems using Psychometric Tests

Fall 2009 - Spring 2011

- Suggested a novel approach to use Psychometric Tests (Luscher Color Test and Myers Briggs Type Indicator Test) to evaluate psychometric index of users
- Cluster nodes with similar scores and in case of a particularly high-frequency zone, we treat these nodes as suspicious and further use CAPTCHAs to remove false positives.

Software Quality Evaluation using Fuzzy Multi-Criteria Approach

Spring 2010- Spring 2011

- Employed fuzzy ratings and weights to software attributes and proposed a comprehensive model for calculating overall software quality based on ISO/IEC 9126 model.
- Tested on multiple university softwares and one industrial application.

Teaching Assistant

Jan - May 2012

Assisted the instructor in teaching the following undergraduate level courses. Duties included sharing of responsibilities for exams, homework assignments, grades and leading computer lab exercises.

- CS/IS 332 Introduction to Database Systems and Application, Spring 2012.

PUBLICATIONS:
CONFERENCES

A. Paul, A. Agrawal, W. Liao and A. Choudhary. "AnonyMine: Mining anonymous social media posts using psycho-lingual and crowd-sourced dictionaries", *Proceedings of the Workshop on Issues of Sentiment Discovery and Opinion Mining at 22nd Annual ACM Conference on Knowledge Discovery and Data Mining, 2016.*

J.Birnholtz, N.A.R. Merola, and **A. Paul**. "Is it Weird to Still Be a Virgin??: Anonymous, Locally Targeted Questions on Facebook Confession Boards", *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015.*

A. Paul, Varuni G., J.S. Challa and Y. Sharma "HADCLEAN: A Hybrid Approach for Data Cleaning Techniques in Data Warehouses", *Proceedings of the IEEE International Conference on Information Retrieval and Knowledge Management(CAMP),Kuala Lumpur, March,2012*

A.Paul*, J.S. Challa, Y. Dada, V. Nerella, P.R. Srivastava "Quantification of Software Quality Parameters using Fuzzy Multi-Criteria Approach," *Proceedings of the IEEE International Conference on Process Automation Control and Computing (PACC) 2011, Coimbatore, July, 2011*

A.Paul*, K Haribabu and C. Hota “**Detecting Sybils in Peer-to-Peer Overlays using Psychometric Analysis Methods**,”, *Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications(AINA)*, Singapore, March 2011

JOURNALS

K Haribabu, C.Hota and **A. Paul** “**GAUR: A Method to Detect Sybil Groups in Peer-to-Peer Overlays**”, *International Journal of Grid and Utility Computing*, 2012 Vol.3 ISSN : 1741-847X
<http://dx.doi.org/10.1504/IJGUC.2012.0477655>

A.Paul*, J.S. Challa, Y.Dada, V.Nerella, P.R. Srivastava and A.P.Singh “**Integrated Software Quality Evaluation: A Fuzzy Multi-Criteria Approach**”, *Journal of Information Processing Systems (JIPS): Korean Information Processing Society, Volume 7, Number 3 (September 2011) ISSN : 1976-913X*.
<http://dx.doi.org/10.3745/JIPS.2011.7.3.473>

* = co-first author

PROFESSIONAL EXPERIENCE

Narus Inc. - A Boeing Company, Sunnyvale, California

Summer Research Intern

Jun 2013 - Sep 2013

- Understanding Collaboration Among Online Advertising and Analytics Services
- Observed multiple 3rd-party services sharing user's private information with each other
- Investigated how these services use means to obfuscate this parameter sharing

BITS Pilani, Rajasthan, India

Project Assistant

Feb-Dec 2011

- Database Administrator(Oracle 11g Server) for the university Information Processing Center: Worked on Data mining and Data Warehousing tools used over Oracle 11g
- Developed a customized version of Moodle 1.9/2.0 over LAMP stack for the university On-Campus Course Management System

National Thermal Power Corporation, Delhi, India

Undergraduate Intern

June-Aug 2011

- Studied water treatment processes and recommended improved water treatment measures
- Prepared research design and surveys to study effects of pollution from the plant as part of the Environment Management Group

FELLOWSHIPS

- **Recipient of Walter P. Murphy Fellowship**, during 1st year of PhD (2012-2013)
- **Recipient of Segal Design Cluster Fellowship**, during 3rd year of PhD (2014-2015)
- **Recipient of Predictive Science and Engineering Design Fellowship**, during 5th year of PhD (2016-2017)
- **Recipient of BITS Pilani Merit-cum-Need Scholarship** during undergraduate study

AWARDS & ACHIEVEMENTS

- Among 10 doctoral students across Northwestern selected for summer-long Research Communication Workshop, 2016
- Most viewed Writer on Quora in following categories: Natural Language Processing, Computational Linguistics, Natural Language Processing Tools, Feature Extraction
- **Best TA award** for recognition of teaching excellency as Teaching Assistant for Database Systems and Applications (BITS Pilani 2012)
- **All India Rank 1** in BITS HDSAT (admission test for graduate programs at BITS Pilani) in Software Systems
- **All India Rank 64** & State Rank 9 in National Science Olympiad among more than half million participants during freshmen year of high-school

SELECTED COURSE PROJECTS

- Developed a Sentiment Analysis Tool to find the most interesting or controversial events at the 2013 Golden Globe Awards from user-Tweets (Python) **Spring 2014**
- Developed a tool which uses Natural Language Processing techniques to find the most interesting or controversial events at the 2013 Golden Globe Awards from user-Tweets (Python) **Winter 2014**
- Developed Sudoku & Othello solver using constraint satisfaction and min-max algorithms using efficient tree-based data structures and algorithms (C++). **Spring 2013.**

- Implemented a fully distributed event detection mechanism by utilizing a Kademlia-based DHT overlay network (Go) **Winter 2013**
- Developed a web application to track a portfolio of a user's stocks. Used data mining techniques to analyze and predict stock and portfolio performance using historical data. (Perl, SQL) **Fall 2012**

SELECTED SIDE PROJECTS

- Developed a tool which creates a recommendation system using ElasticSearch(Lucene) for shopping based on "I just bought" Amazon tweets of users (Python) **Winter 2014**
- Developed a real-time tool starts an alarm when a designated bus is 'x' (customizable) min away from the closest bus stop by scraping CTA bus tracker webpage (Python). **Fall 2014.**
- Developed a web-automation & scraping tool which collects past news articles from the web. Used OCR recognition for getting the text from old articles (Python). **Spring 2014.**

COMPUTER SKILLS

Proficient: Python, C, C++, Java, PHP, Perl, Selenium, LAMP, Go, VB, MySQL, shell scripting, weka
Familiar: R, Gephi, Javascript, html,css, MATLAB, Hadoop, Mahout, OpenMP, MPI, .NET, Oracle, MS-SQL Server

LEADERSHIP

- President, Northwestern University Cricket Club
- Treasurer and Founding Member, Northwestern Toastmasters Club
- Mentor, Brave Initiatives (Teaching girls to code: <http://www.braveinitiatives.com/>)
- Social Media Chair, Northwestern Tango Club
- STEM Liaison, Northwestern Ethnic Students Group