

# Denzil Correa

Bayer AG

✉ [denzil.correa@bayer.com](mailto:denzil.correa@bayer.com)

🌐 [www.correa.in](http://www.correa.in)

Data Scientist

## Education

- 2009–2014 Ph.D Computer Science, IIIT-Delhi, India
- 2012 Masters in Computer Science, IIIT-Delhi, India
- 2007 Bachelors in Computer Science, University of Mumbai, India

## PhD thesis

- title *Content Quality in Web 2.0 Services : Analysis, Detection and Systems*
- supervisors Prof. Ashish Sureka
- description This thesis explores quality management and enhancement of user generated content quality on crowd sourcing systems. We discover unique, significant characteristics of low quality content and propose feasible automated techniques to detect such content. We also build systems to help Internet users post better quality content. We use qualitative and quantitative methods to show the efficacy of our systems.

## Master thesis

- title *WhACKY! What Anyone Could Know About You from Twitter*
- supervisor Prof. Ashish Sureka
- description This thesis explores the idea of a behavioral context based threat attack to reveal personally identifiable information about users on social media. The thesis outlines an attack based on unintentional social cues and details the implications on privacy in social media.

## Awards and Achievements

- 2015 Indo-German Max Planck Post Doctoral Fellowship
- 2013 1st Prize, Govt. of India Data Hackathon (Visualization)
- 2010-2012 NIXI Travel Fellowship [*Acceptance Rate : 4%*]
- 2010-2014 Information Technology Ph.D Fellowship, Govt. of India

## Skills

- Research** Big Data Analytics, Data Science
- Programming** Python, MySQL, NLTK, scikit-learn, NumPy, Matlab, R, Java

---

## Patents

- title *System and Method for Verifying Credentials*  
inventors **Denzil Correa** and Ashish Sureka  
full text <http://www.google.com/patents/US20130275753>  
description The patent outlines a system to authenticate educational credentials in a privacy preserving protocol.

---

## Experience

### Industry

- 2016- **Data Scientist**, *Bayer AG*, Germany
- Machine Translation via Deep Learning
  - Mining Social Media for Drug Side Effects
- 2007-2009 **Research Engineer**, *Infosys Research*, India
- Tabular Data Extraction from PDF documents
  - Automatic Index Generation System for Journals
  - Automated Patent Document Classification System

### Academia

- 2015 **Post Doctoral Fellow**, *Max Planck Institute for Software Systems*, Germany  
Supervisor Prof. Krishna P. Gummadi
- Study of Cultural Attitudes and Perceptions via Social Media
- 2014 **Intern**, *Max Planck Institute for Software Systems*, Germany  
Supervisor Prof. Krishna P. Gummadi
- Characterization of Anonymous Social Media Content
- 2012 **Intern**, *Singapore Management University*, Singapore  
Supervisor Prof. Ee Peng Lim
- Framework to Revive Dormant Friendships in Social Media
- 2010 **Intern**, *IBM Research*, India
- Framework to Detect Spam on Twitter
- 2009-2014 **Ph.D Student**, *IIT-Delhi*, India  
Supervisor Prof. Ashish Sureka
- Low Quality Content Analysis and Detection on Community QnA sites
  - Productivity Enhancement for Software Maintenance Professionals
  - Social Behavioral Context based Privacy Threat Attack on Social Media
  - Topic-Based Community Detection
  - Hate Speech Characterization on Social Media
  - Sentiment Analysis and Opinion Mining on Product Reviews
- 2006 **Intern**, *IIT-Bombay*, India
- Query by Humming - Musical Information Retrieval System

---

## Journal Publications

- 2016 • Agarwal, Swati, Nitish Mittal, Rohan Katyal, Ashish Sureka, and **Denzil Correa**. "Women in Computer Science Research: What is the Bibliography Data Telling Us?" In: *SIGCAS Comput. Soc.* 46.1 (Mar. 2016), pp. 7–19.

---

## Conference Publications

- 2016 • Silva, Leandro, Mainack Mondal, **Denzil Correa**, Fabrício Benevenuto, and Ingmar Weber. "Analyzing the Targets of Hate in Online Social Media". In: *Tenth International AAAI Conference on Web and Social Media*. 2016.

- 2015 • **Denzil Correa**, Leandro Araujo Silva, Mainack Mondal, Fabricio Benevenuto, and Krishna P Gummadi. "The Many Shades of Anonymity: Characterizing Anonymous Social Media Content". In: *Ninth International AAAI Conference on Web and Social Media*. 2015.

- 2014 • **Denzil Correa** and Ashish Sureka. "Chaff from the wheat: Characterization and modeling of deleted questions on stack overflow". In: *Proceedings of the 23rd international conference on World wide web*. ACM. 2014, pp. 631–642.

- 2013 • Lim, Ee-Peng, **Denzil Correa**, David Lo, Michael Finegold, and Feida Zhu. "Reviving Dormant Ties in an Online Social Network Experiment". In: *Seventh International AAAI Conference on Weblogs and Social Media*. 2013.

- 2013 • **Denzil Correa**, S. Lal, A. Saini, and A. Sureka. "Samekana: A Browser Extension for Including Relevant Web Links in Issue Tracking System Discussion Forum". In: *Software Engineering Conference (APSEC), 2013 20th Asia-Pacific*. Vol. 1. 2013, pp. 25–33.

- 2013 • **Denzil Correa** and Ashish Sureka. "Fit or Unfit: Analysis and Prediction of 'Closed Questions' on Stack Overflow". In: *Proceedings of the first ACM conference on Online social networks*. ACM. 2013, pp. 201–212.

- 2013 • **Denzil Correa** and Ashish Sureka. "Integrating issue tracking systems with community-based question and answering websites". In: *Software Engineering Conference (ASWEC), 2013 22nd Australian*. IEEE. 2013, pp. 88–96.

- 2012 • **Denzil Correa**, Ashish Sureka, and Mayank Pundir. "iTop: interaction based topic centric community discovery on twitter". In: *Proceedings of the 5th Ph. D. workshop on Information and knowledge*. ACM. 2012, pp. 51–58.

2012

**Denzil Correa**, Ashish Sureka, and Raghav Sethi. "WhACKY!-What anyone could know about you from Twitter". In: *Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on*. IEEE. 2012, pp. 43–50.

2011

**Denzil Correa** and Ashish Sureka. "Mining tweets for tag recommendation on social media". In: *Proceedings of the 3rd international workshop on Search and mining user-generated contents*. ACM. 2011, pp. 69–76.

2010

Sureka, Ashish, Vikram Goyal, **Denzil Correa**, and Anirban Mondal. "Generating Domain-Specific Ontology from Common-Sense Semantic Network for Target-Specific Sentiment Analysis". In: *computer and information science, Artificial Intelligence paper*. 2010.

2009

Sureka, Ashish, Vikram Goyal, **Denzil Correa**, and Anirban Mondal. "Polarity Classification of Subjective Words Using Common-Sense Knowledge-Base". In: *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*. Springer Berlin Heidelberg, 2009, pp. 486–493.

---

## Academic Service

### Journal Reviewer

- Transactions on Knowledge Discovery from Data (TKDD)
- Behaviour & Information Technology
- International Journal of Human-Computer Studies (IJHCS)
- Computer Communications - Special Issue on Online Social Networks (ComCom)

### Conference Reviewer

- Conference on Computer Supported Cooperative Work (CSCW)
- Computer Human Interaction (CHI)
- Asia Pacific Software Engineering Conference (APSEC)
- Indian Software Engineering Conference (ISEC)
- International Symposium on Software Reliability Engineering (ISSRE)