MD IMRUL KAYES

3715 Jefferson Commons Dr., 301, Tampa, FL 33613
imrul@mail.usf.edu, (813) 405 6320
http://www.cse.usf.edu/~imrul/

SUMMARY

- 3+ years of experience in software development and quality.
- 4+ years of experience in data science: mining large, complex networks ("Big Data"), such as online social, community Q/A and blogging using data mining techniques; studying user behavior based on psychology and sociology theories.
- PhD and MS degrees in Computer Science & Engineering.
- Big Data summer school scholarships and grants from Google, ACM and SoftwarePeople.
- Worked for Yahoo, DSG Labs, SoftwarePeople, Delta Life, and Binary Solutions.

EDUCATION

PhD, Computer Science & Engineering — **University of South Florida** — **2015**
- **Dissertation**: Content abuse and privacy concerns in online social networks.

MS, Computer Science (GPA: 3.85) — **University of South Florida** — **2013**

BSc, Computer Sci. & Engg. (GPA: 3.61) — **Bangladesh Uni. of Engg. and Tech.** — **2009**

WORK EXPERIENCE

Yahoo Labs, Barcelona — **Research Intern** — **September'13 - December'13**
- Detected content abusers in Yahoo Answers.
- Used Hadoop, Pig Latin, Python, R, and Awk to build a data pipeline to transform raw data for analysis.
- Analyzed Yahoo Answers data (more than one year recorded activity logs of 1.5 million users), created exploratory graphs, conducted statistical tests and built machine learning models.

Distributed Systems Lab, Tampa, FL — **Research Assistant** — **May'12 - Present**
- Designed and Implemented *Aegis* privacy framework that enforces privacy as contextual integrity in social ecosystems.
- Analyzed user retention in community blogs and built machine learning models to predict retention.
- Proposed a method to identify influential bloggers based on network centralities and detected influential bloggers.
- Analyzed privacy concerns and user behavior in CQAs and built machine learning models to predict privacy.
- Showed that cross-cultural variations exist in CQAs analyzing user activity logs.

SoftwarePeople, Copenhagen — **Software Engineer** — **March'10 - July'11**
- Worked as a Scrum team member for the development of Agile-based software.
- Used C#, .NET, MS SQL server to build web-based business applications and service oriented systems.
- Researched on software quality and testing, wrote research papers, awarded a travel grant to present the paper.

Delta Life, Dhaka — **Software Engineer** — **September'09 - December'09**
- Developed "Xfer" using C# and Oracle, which synchronizes data transfer from remote servers to the central server.
- Worked on the development of the official website of Delta Life Insurance Co. Limited.

Binary Solutions, Dhaka — **Software Developer** — **November'08 - August'09**
- Developed "EasyShare" using C# and MySQL. EasyShare is a share market analyzer software, which collects data from the Dhaka Stock Exchange server and shows comparative market situation.

TECHNICAL SKILLS

Languages	Python, C, C++, Java, C#, ASP.NET, R, PL/SQL, HTML
Database	MongoDB, Oracle, MS Access, SQL Server
Distributed Data Processing	Hadoop (Map Reduce), Pig Latin
Data Mining	Weka, R
Social Network Analysis Tools	NetworkX, JUNG, SNAP, Gephi
Applications	OmniGraffle, LaTeX, Gnuplot, Git

HONORS AND AWARDS

- ACM travel grant to the 26th conference on Hypertext and Social Media, Cyprus, 2015.
- San Diego Supercomputer center scholarship to the Big Data summer school, San Diego, 2014.
- Google travel grant to the 4th SocInfo conference, Switzerland, 2012.
- SoftwarePeople travel grant to the IEEE ICECT conference, India, 2011.

SELECTED PROJECTS

Retention in Community Blogs.
Used machine learning techniques to determine what factors are associated with users' continued participation.

- Wrote a multi-threaded and error-resistant crawler in Python to scrape an Ajax-based blogging platform, "Blogster".
- Collected a sample of blogger profiles (contributed 91% blogs) and stored in a No-SQL database (MongoDB).
- Analyzed retention using NetwrokX and R: predicted top retained users with 93.62% accuracy.

Aegis: A Semantic Implementation of Privacy as Contextual Integrity in Social Ecosystems.
Modeled user privacy as contextual integrity using semantic web tools.

- Proposed an ontology-based data model to capture users aggregated social data from diverse sources.
- Used semantic tools (RDF/SPARQL) to generate default privacy policies based on contextual integrity.
- Designed an architecture and implemented the privacy model using Java and Jena.

Content Abusers in Community Question Answering.
Used user-contributed rule violation reports to characterize and detect content abusers in Yahoo Answers.

- Used Hadoop, Pig, Python, and R to analyze one year recorded activity logs of 1.5 million Yahoo Answers users.
- Showed that users are good at flagging content and flagged content are removed quickly.
- Showed that moderate deviance is not necessarily bad, but extreme deviant users are likely to be suspended.
- Built machine learning models that were able to detect content abusers with 83% accuracy.

Cultures in Community Question Answering.
Investigated the influence of national culture on people's online questioning and answering behavior.

- Analyzed a sample of 200K users in Yahoo Answers from 67 countries using a number of cultural factors extracted from Geert Hofstede's cultural dimensions and Robert Levine's Pace of Life.
- Showed cultural variations in the predictability of activities, contribution behavior, privacy, and power distance.

Prioritizing Test Cases for Regression Testing.
Prioritized test cases for regression testing based on the dependency network of faults.

- Modeled software fault dependencies as a directed graph and identified leading faults to prioritize test cases.
- Developed a subject software "Tarantula" (19,390 lines of code) using C#, Python and compared the effectiveness of the proposed approach with traditional techniques.

SELECTED PEER-REVIEWED PUBLICATIONS

[1] **I. Kayes,** N. Kourtellis, D. Quercia, A. Iamnitchi, F. Bonchi, The Social World of Content Abusers in Community Question Answering. *ACM 24th World Wide Web conference (WWW'15), Italy, 2015.* (Acceptance rate: 14.10%)
[2] **I. Kayes,** J. Chakareski, Retention in Online Blogging: A Case Study of the Blogster Community. *IEEE Transactions on Computational Social Systems.(In press)*
[3] **I. Kayes,** N. Kourtellis, F. Bonchi, A. Iamnitchi, Privacy Concerns vs. User Behavior in Community Question Answering. *IEEE/ACM Conf. on Advances in Social Networks Analysis and Mining (ASONAM'15), France, 2015.*
[4] **I. Kayes,** N. Kourtellis, D. Quercia, A. Iamnitchi, F. Bonchi, Cultures in Community Question Answering. *ACM 26th Conference on Hypertext and Social Media (HT'15), Cyprus, 2015.*
[5] **I. Kayes,** S. Islam, J. Chakareski, The Network of Faults: A Complex Network Approach to Prioritize Test cases for Regression Testing. *Springer Innovations in Systems and Software Engineering Journal (2015).*
[6] **I. Kayes,** X. Zuo, D. Wang, J. Chakareski , To Blog or Not to Blog: Characterizing and Predicting Retention in Community Blogs, *ACM 7th Social Computing Conference (SocialCom'14), China, 2014.* (Acceptance rate: 16.6%)
[7] **I. Kayes,** A. Iamnitchi, Aegis: A Semantic Implementation of Privacy as Contextual Integrity in Social Ecosystems, *IEEE 11th Conference on Privacy, Security and Trust, Spain, 2013.* (Acceptance rate: 28.76%)
[8] **I. Kayes,** X. Qian, J. Skvoretz, A. Iamnitchi, How Influential are You: Detecting Influential Bloggers in a Blogging Community, *4th International Conference on Social Informatics, Switzerland, 2012.* (Acceptance rate: 25.9%)