

# Abu Awal Md Shoeb

- ✉ [abu.shoeb@rutgers.edu](mailto:abu.shoeb@rutgers.edu)
- 🐦 [@abushoeb](https://twitter.com/abushoeb)
- 🌐 <https://shoeb.info/>
- 🌐 <http://www.linkedin.com/in/abushoeb/>



## About

---

- 📌 A passionate, innovative technologist with research, data science and leadership experience
- 📌 Current research involves automatic annotation, segmentation, and classification of clinical notes to improve the existing 3M OneNLU pipeline

## Education

---

- 2015 – 2021 **Ph.D. in Computer Science**  
Rutgers University, New Brunswick, New Jersey, USA  
Dissertation: *Computational Methods to Understand the Association between Emojis and Emotions* [PDF]  
Advisor: Gerard de Melo
- 2013 – 2015 **M.Sc. in Computer and Information Sciences**  
University of Alabama at Birmingham (UAB), Birmingham, Alabama, USA  
Research Area: Cloud Computing, Cyber Security  
Advisor: Ragib Hasan
- 2003 – 2008 **B.Sc. in Computer Science and Engineering**  
Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh  
Thesis: *Grid computing*  
Advisor: Mohammad Khalad Hasan



## Employment History

---

- 2021 – Present **Machine Learning Research Scientist - Remote**  
3M Health Information Systems  
Saint Paul, Minnesota, USA
- 2015 – 2021 **Teaching Assistant**  
Dept. of Computer Science, Rutgers University
- 2015 – 2020 **Graduate Mentor**  
Douglass DIMACS Computing Corps (DDCC), Rutgers University  
- Provided guidance and support to undergraduate women students in STEM to design and carry out a fun and interactive group activities for middle school girls to discover the excitement of computing and programming
- Summer 2018 & 2017 **Summer Research Intern**  
Nokia Bell Labs  
Murry Hill, New Jersey, USA






## Employment History (continued)

---

- Summer 2015 & 2014  **Software Engineering Intern**  
MedSnap LLC, Birmingham, Alabama, USA  
- Researched drug-related data and images and developed a visualization tool to help to identify counterfeit medication  
- Technologies: Vagrant, PHP, Amazon EC2, d3, JavaScript, MSSQL
- 2013 – 2015  **Graduate Research Assistant**  
SeCuRE and Trustworthy Computing Lab (SECRETLab)  
University of Alabama at Birmingham (UAB), Birmingham, Alabama  
Research Tasks:  
- Spam Email Campaign Detection using Data Mining  
- Performance Analysis of Amazon EC2 Instances  
- Sockpuppet detection in Wikipedia using Machine Learning Techniques.  
- Forensics Analysis of a Hard Drive using EnCase.  
- Enhanced the performance of a web application for medication safety and developed a visualization tool including the migration to the Amazon cloud.
- Feb 2013 – Jun 2013  **External IT Consultant**  
Dhaka Stock Exchange Ltd., Dhaka, Bangladesh  
- Member of an investigative team to identify the origin of the problem of new MSA Plus Trading software launched by Dhaka Stock Exchange Ltd.
- 2013 – 2019  **Founder and CEO**  
Criss Cross Computers, Sylhet, Bangladesh  
- Developed and implemented the first online service for students at SUST to make payments and withdraw transcripts [Links]
- 2009 – 2010  **Project Manager**  
Center for Research Testing and Consultancy (CRTC), CSE, SUST  
- Introduced mobile SMS based registration and payment systems for undergraduate admission test first time in Bangladesh  
- Team Size 11+, End users ~2.3M
- 2008 – 2009  **e-Service Programmer**  
Access to Information (A2I) program, Prime Minister's Office  
United Nations Development Programme (UNDP), Dhaka, Bangladesh  
- Designed and implemented framework for ICT services for different ministries of Bangladesh Architectural and database design of freedom fighters web application of 1971 liberation war of Bangladesh

## Teaching Experience





---

- 2015 – 2021  **Rutgers University, New Brunswick, NJ**  
Teaching Assistant, Dept. of Computer Science  
Courses:  
- CS 211 Computer Architecture (Overall student rating 4.53/5)  
- CS 214 System Programming (Overall student rating 4.25/5)  
- CS 314 Principles of Programming Languages (Overall student rating 4.22/5)  
- CS 439 Introduction to Data Science (Overall student rating 4.19/5)  
- CS 543 Massive Data Storage and Retrieval (Overall student rating 4.09/5)
- 2009 – 2013  **Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh**  
Assistant Professor, Dept. of Computer Science & Engineering (CSE)  
Courses:  
- CSE 223 Electronic Communication (Class size 60)  
- CSE 233 Object Oriented Programming Java (Class size 85)  
- CSE 333 Database Management System (Class size 120)  
- CSE 431 Computer Networking (Class size 75)  
- CSE 445 Web Engineering (Class size 80)
-  **Center for Research Testing and Consultancy (CRTC), CSE, SUST, Bangladesh**  
Adjunct Faculty  
Courses:  
- Basic Web design using CSS, HTML, JavaScript, PHP  
- Object Oriented Programming Java  
- Structured Programming Language C
- 2010 – 2013  **Leading University, Sylhet, Bangladesh**  
Adjunct Faculty  
Courses:  
- Object Oriented Programming Java  
- Structured Programming Language C  
- Relational Database Management Systems
- 2011 – 2011  **Metropolitan University, Sylhet, Bangladesh**  
Adjunct Faculty  
Courses:  
- Object Oriented Programming Java

## Research Publications

---

### Conference Proceedings

-  Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., ... Wu, Z. (2022). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. (Vol. 2206.04615). arXiv: 2206.04615 [cs.CL]. Retrieved from  <https://arxiv.org/abs/2206.04615>
-  Shoeb, A. A. M., & de Melo, G. (2021). Assessing emoji use in modern text processing tools. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 1379–1388).  doi:10.18653/v1/2021.acl-long.110

- 3 Shoeb, A. A. M., & de Melo, G. (2020). EmoTag1200: Understanding the association between emojis and emotions. In *Proceedings of the 2020 conference on empirical methods in natural language processing (emnlp)* (pp. 8957–8967). Online: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/2020.emnlp-main.720>
- 4 Shoeb, A. A. M., Raji, S., & de Melo, G. (2019). EmoTag – towards an emotion-based analysis of emojis. In *Proceedings of the international conference on recent advances in natural language processing (ranlp 2019)* (pp. 1094–1103). doi:10.26615/978-954-452-056-4\_126
- 5 Shoeb, A. A. M., Mukhopadhyay, D., Al Noor, S., Sprague, A., & Warner, G. (2015). Spam campaign cluster detection using redirected urls and randomized sub-domains, Academy of Science and Engineering, USA.
- 6 Shoeb, A. A. M., Hasan, R., Haque, M., & Hu, M. (2014). A comparative study on i/o performance between compute and storage optimized instances of amazon ec2. In *2014 IEEE 7th international conference on cloud computing* (pp. 970–971). doi:10.1109/CLOUD.2014.146
- 7 Shoeb, A. A. M. (2011). Economic benefits of sms based mobile application in university admission. In *The national conference on contemporary issues in economics (nccie)*.
- 8 Shoeb, A. A. M., Hussain, A., Bikas, M. A. N., & Hasan, M. K. (2009). An extended algorithm to enhance the performance of the gridbus broker with data restoring technique. In *2009 international conference on computer engineering and technology* (Vol. 1, pp. 371–375). IEEE.
- 9 Bikas, M. A. N., Hussain, A., Shoeb, A. A. M., Hasan, M. K., & Rabbi, M. F. (2008). File based grid thread implementation in the .net-based alchemi framework. In *2008 IEEE international multitopic conference* (pp. 468–472). IEEE.
- 10 Hussain, A., Shoeb, A. A. M., Bikas, M. A. N., & Hasan, M. K. (2008a). File based grid thread implementation in the .net based alchemi framework. In *IEEE xplore, multitopic conference inmic*.
- 11 Hussain, A., Shoeb, A. A. M., Bikas, M. A. N., & Hasan, M. K. (2008b). Runtime thread rescheduling: An extended scheduling algorithm to enhance the performance of the gridbus broker. In *2008 11th international conference on computer and information technology* (pp. 80–87). IEEE.

## Journal Articles

- 1 Al Mumin, M. A., Shoeb, A. A. M., Selim, M. R., & Iqbal, M. Z. (2014). Sumono: A representative modern bengali corpus. *SUST Journal of Science and Technology*, 21(1), 78–86.
- 2 Al Mumin, M. A., Shoeb, A. A. M., Selim, M. R., & Iqbal, M. Z. (2012). Supara: A balanced english-bengali parallel corpus. *SUST Journal of Science and Technology*, 16(2), 46–51.
- 3 Chowdhury, S. S., & Shoeb, A. A. M. (2012). Performance analysis of mpi (mpi4py) on diskless cluster environment in ubuntu. *International Journal of Computer Applications*, 60(14).
- 4 Shamanta, D., Rahmi, M., & Shoeb, A. A. M. (2012). Focused web crawling: An architecture for crawling of country based financial data. *International Journal of Advanced Research in Computer Science*, 3(5).

## Posters

- 1 Shoeb, A. A. M., & Raji, S. (2017). Query social media (twitter) with hashtag and analyze them by using microsoft cognitive services.
- 2 Shoeb, A. A. M., & Imtiaz, H. (2016). Utilizing hashtag in opinion mining and sentiment analysis for customers survey.


## Technical Reports

- 1 Shoeb, A. A. M. (2018a). *Deep learning job scheduling under kubernetes (k8) cluster*.





- 2 Shoeb, A. A. M. (2018b). *Is private browsing in modern web browsers really private?*
- 3 Shoeb, A. A. M. (2015). *Forensics analysis of a hard drive using encase*. Retrieved from <http://shoeb.info>

## Professional Activities









### Conference Program Committee Member

2021  SocialNLP in conjunction with NAACL 2021 [Sites]








### Reviewer

ACL  2022, 2021  
 CoNLL  2022, 2021  
 EMNLP  2022, 2021  
 Journal  Computational Intelligence Journal, 2020

### Presentations and Invited Talks

November 2020  Emojis: Modern Tools for Emotions, ZipRecruiter  
 March 2020  Emojis: Modern Tools for Emotions, CS Open House, Rutgers University  
 November 2019  Pitched a business idea in Institute for Biomedical Entrepreneurship program, Boston, Massachusetts  
 September 2019  Presented my research work “EmoTag – Towards an Emotion-Based Analysis of Emojis”, in RANLP Conference, Varna, Bulgaria  
 December 2014  Presented my research work “Spam Campaign Cluster Detection Using Redirected URLs and Randomized Sub-Domains” in Social Informatics 2014, December 14-16, 2014, Cambridge, MA, USA [Video]  
 March 2010  “Challenges of a Paperless Admission System”, SQABD Lightning Talks 4, March 2010, Dhaka, Bangladesh [Video]  
 January 2010  “An Extended Algorithm to Enhance the Performance of the Gridbus Broker with Data Restoring Technique”, International Conference on Computer Engineering and Technology, Singapore 2010  
 2009  Presented a paper on “Digital Bangladesh and What to do” as a keynote speaker in a national seminar in 2009, Sylhet, Bangladesh

### Awards and Achievements

2020  Nomination for **Chancellor Leadership Award**, Rutgers University, Spring 2020  
 2019  **Rutgers School of Arts and Sciences Grants** to attend Biomedical Innovation and Entrepreneurship Certificate Program in Boston, Massachusetts, November 2019  
 Dept. **Travel Grants** to attend RANLP Conference, Varna, Bulgaria  
 Invitation to participate **Rutgers I-Corps Cohort 5**, Summer 2019  
 2015-21  **Graduate Teaching Assistantship**, Dept. of Computer Science, Rutgers University (Fall 2015 - Spring 2021)  
 2016-18  **TA-GA Professional Development Fund**, Rutgers University, Summer 2016, 2017, 2018  
 2015-20  **Graduate Mentor**, Douglass DIMACS Computing Corps, Rutgers University (November 2015 - 2020)

## Professional Activities (continued)

---

- 2014 ■ **Best Paper Award** in ASE International Conference on Social Informatics 2014, Harvard University, Cambridge, MA, USA (Dec 14-16, 2014)
- 2014 ■ **Dept. Travel Grant from University of Alabama at Birmingham** to attend ASE International Conference on Social Informatics 2014, Harvard University, Cambridge, MA, USA (Dec 14-16, 2014)
- 2013-15 ■ **Graduate Research Assistantship**, University of Alabama at Birmingham (Fall 2013 – Spring 2015)
- 2010 ■ **Emerging Young Talents** (for the year 2010 by the Prothom Alo, the most popular and leading daily newspaper in Bangladesh) for introducing Mobile SMS Based Registration System for University Admission Tests [Link]
- 2010 ■ **mBillionth South Asian Award** for introducing Mobile SMS Based Registration System for University Admission Tests, July 2010 [Link]

## Certification

- 2019 ■ **Biomedical Innovation and Entrepreneurship** Certificate Program, IBE, Boston, Massachusetts, November 2019
- 2019 ■ Graduated from **Y Combinator Startup School**, September 2019

## Other Services

- 2020 ■ **Judge for HackHERS**, a women-centered 24-hour hack-a-thon, Rutgers University, February 2020

## Volunteer Services

---

- Fall 2018 ■ **Discussion Leader**, International TA Orientation, Rutgers University
- 2016 – 2017 ■ **President**, Graduate Student Association (GSA), Rutgers University
- 2013 – 2017 ■ **Elected Syndicate Member**, Shahjalal University of Science and Technology, Sylhet, Bangladesh
- 2010 ■ **Adviser**, SUST Prothom-Alo Bondhushava
- 2007 – present ■ **Core Member**, Bangladesh Open Source Network (BdOSN)
- 2007 ■ **Volunteer**, Bangladesh National Mathematical Olympiad
- 2006 – 2007 ■ **General Secretary**, CSE Society, SUST

## Skills

---

- Languages ■ Strong reading, writing and speaking competencies for English, Mandarin Chinese, Bahasa Malaysia.
- Coding ■ Java, PHP, Python, R, SQL, XML/XSL,  $\LaTeX$ , ...
- Databases ■ MySQL, PostgreSQL, HSQL, SQLite, LotusNotes.
- Web Dev ■ HTML, CSS, JavaScript, Apache Web Server, Tomcat Web Server.
- Misc. ■ Academic research, teaching, training, consultation,  $\LaTeX$  typesetting and publishing.

## References

---

Available on request