Sanchit Alekh

Big Data Engineer, Data Reply GmbH

Webpage: salekh.github.io/about

Github: https://github.com/salekh

+49 152 3627 6687

sanchit.alekh@gmail.com

linkedin.com/in/sanchitalekh

Scholar: https://goo.gl/hQJ9oz

Education

- **RWTH Aachen University** Aachen, Germany
  MSc. - Software Systems Engineering *2016 - 2018*
  - GPA: *1,3 (Sehr gut/Very Good)* on the German Academic Grading Scale
  - Area of Specialization: Data and Information Management
  - Master Thesis at AUDI AG on the topic of data quality management for enterprise data lakes. Available at `shorturl.at/iyAY2`

- **Indian Institute of Information Technology** Allahabad, India
  B.Tech (Honours) - Information Technology *2012 - 2016*
  - GPA: *9.09/10.0*
  - Received a *Merit Incentive Scholarship* from Government of India
  - Pursued Bachelor Thesis on the topic *Author Disambiguation for Medical Patents* at RWTH Aachen University, Germany

- **Delhi Public School** Patna, India
  Higher Secondary Schooling (Abitur), Score: 93.6 % *2010 - 2012*

Skills

- **Programming Languages:** Java, Python, Scala, Typescript with Node.js, LaTeX

- **Interest Areas:** Data Platforms, Conversational Systems, Data/Text mining algorithms, Statistical algorithms, Privacy Threat Modeling of Software Systems, Ontology Engineering and Linked Data, Computer Vision

- **Technologies:** Cloud Platforms and Technologies, Distributed Frameworks (e.g. Apache Spark, Apache Kafka), MQTT/OPC-UA, Databases (SQL and NoSQL), Linux, Git, Medical Informatics Standards

Research and Work Experience

- **Data Reply GmbH** Munich, Germany
  Big Data Engineer *Mar 2019 - current*
  - **Work Summary:** Currently working as a Big Data Engineer, working with clients across Germany and Europe. Focused on building automated, large-scale data solutions on the cloud.

  - **Current Project:** Developing an enterprise conversational platform for the next-generation workplace, at a leading automotive company based in Munich. The platform is built on Microsoft Azure, and leverages cutting-edge NLU capabilities .

- **Areas Spanned:** Public Cloud, Data Lake, Semi-structured Databases, Data Mining, Big Data, Distributed Systems, Cloud-native, Serverless Deployments

- **Audi AG**   Heilbronn, Germany
  Data Scientist Intern   *Jan 2018 - Jan 2019*
    - **Work Summary:** Worked at the TechHub: Data Driven Production group, which leverages big data technologies to solve data-driven use-cases in automotive production.
    - **Areas Spanned:** Data Lake, Semi-structured Databases, Data Mining, Big Data, Distributed Systems, Internet of Things, Industry 4.0

- **Informatik 5, RWTH Aachen University**   Aachen, Germany
  Student Assistant   *Jan 2016 - Jan 2018*
    - **Work Summary:** Worked under PD Dr. Christoph Quix as a part of the mi-Mappa project (`dbis.rwth-aachen.de/cms/projects/mi-mappa`), which aimed to identify suitable actors for complex innovation in medical science. Bachelor Thesis was also a part of this project.
    - **Areas Spanned:** Medical Text Analysis, Information Extraction, Text Mining, High-Performance Computing

- **Knowledge Mining and Assessment Group, TU Darmstadt**   Darmstadt, Germany
  Student Assistant   *May 2015 - Jul 2015*
    - **Work Summary:** Under Prof. Dr. Ulf Brefeld, worked on computing a confidence measure on reliability of data on crowd-sourced encyclopedia and commercial encyclopediae
    - **Areas Spanned:** Python, MySQL, Information Extraction, Text Mining, Conditional Random Fields, Wikipedia API

- **University of Wisconsin**   Milwaukee, USA
  Research Intern   *May 2014 - Jul 2014*
    - **Work Summary:** Analyzed and optimized current algorithms for anonymizing personal information in Medical Health Data. Focused on Mu-Argus for structured and MIST for unstructured data. Developed a software application for anonymization of DICOM and HL7 image data
    - **Areas Spanned:** Java, HIPAA, Text Mining

Selected Publications

1. **An Integrated Ontology-based Approach for Patent Classification in Medical Engineering**
   12th International Conference on Data Integration in Life Science, November 2017, Luxembourg.

2. **Ontology Matching for Patent Classification**
   The Twelfth International Workshop on Ontology Matching, October 2017, Vienna, Austria.

3. **Combining Keystroke Dynamics and Face Recognition for User Verification**
   8th IEEE International Conference on Computational Science and Engineering, October 2015, Porto, Portugal.

4. **Ontology-based classification and Analysis of Smart-City Events**
   International Conference on Computational Techniques in Information and Communication Technologies, March 2016, New Delhi, India.

Languages

1. **Hindi** *Mothertongue*

2. **English** *Native-level Fluency*, TOEFL iBT Score: 118/120

3. **German Intermediate Fluency** TELC B2, Grade: 2,0

References

1. **Patrick Bornikoel, Associate Partner, Data Reply GmbH**
   p.bornikoel@reply.de

2. **Prof. Dr. Christoph Quix, Fraunhofer FIT / Hochschule Niederrhein**
   christoph.quix@fit.fraunhofer.de, Telephone: +49 2241 14-3725

3. **Christoph Ring, Audi AG (now at Microsoft Deutschland)**
   de.linkedin.com/in/christoph-ring-65177799 [1]

---

[1]CV updated on 25. June 2020