

Sanchit Alekh

Big Data Consultant bei Data Reply GmbH, München
Webpage: [salekh.github.io/about](https://github.com/salekh)
Github: <https://github.com/salekh>

+49 152 3627 6687
sanchit.alekh@gmail.com
[linkedin.com/in/sanchitalekh](https://www.linkedin.com/in/sanchitalekh)
Scholar: <https://orcid.org/0000-0001-9149-9149>

Bildungsweg

- **RWTH Aachen** Aachen, Deutschland
MSc. - Software Systems Engineering 2016 - 2018
 - Note: 1,3
 - Schwerpunkt: Daten und Informationsmanagement
 - Voraussichtlicher Abschluss: Dezember 2018
- **Indian Institute of Information Technology** Allahabad, Indien
BTech. mit Auszeichnung - Informationstechnologie 2012 - 2016
 - Note: 9.09/10.0
 - Stipendiat des *Merit Incentive Scholarships* der indischen Regierung
 - Bachelorarbeit zum Thema *Author Disambiguation for Medical Patents* an der RWTH Aachen
- **Delhi Public School** Patna, Indien
Abitur, Note: 93.6 % 2010 - 2012

IT-Fähigkeiten

- **Programmierkenntnisse:** Java, Python, Scala, C++, MATLAB, Latex
- **Themenschwerpunkte:** Machinelles Lernen, Daten/Text Mining Algorithmen, Parallel Programmierung, Statistische Algorithmen, Computer Vision Algorithmen, Datenschutz-Bedrohungsmodellierung von Softwaresystemen, Ontologietechnik und Linked Data
- **Technologien:** Spring, MQTT, Hadoop Big Data Stack, Numerical Libraries, Theano, Web Frameworks, Datenbanken (SQL und NoSQL), Linux, Git, Medical Informatics Standards

Forschungs & Arbeitserfahrung

- **Data Reply GmbH** München, Deutschland
Big Data Consultant März 2019 - jetzt
 - **Zusammenfassung:** Derzeit arbeite ich als Big Data Engineer mit Kunden in ganz Deutschland und Europa und konzentriere mich auf die Erstellung automatisierter, umfangreicher Datenlösungen in der Cloud.
 - **Benutzte Technologien:** Öffentliche Cloud, Data Lake, Semi-strukturierte Datenbanken, Data Mining, Big Data, Verteilte Systeme, Cloud-native, Serverlose Bereitstellungen
- **Audi AG** Heilbronn, Deutschland
Praktikant und Masterarbeit im Bereich Data Science Jan 2018 - jetzt

- **Zusammenfassung:** Arbeite bei der TechHub: Data Driven Production, die darauf abzielt, Big-Data-Technologien zu nutzen, um die Produktion automatisierter und effizienter zu gestalten
- **Benutzte Technologien:** Data-Lake, Semi-structured Datenbanken, Data Mining, Big-Data

- **Informatik 5, RWTH Aachen** Aachen, Deutschland
Hilfswissenschaftler Jan 2016 - Jan 2018

- **Zusammenfassung:** Teil des mi-Mappa Projekts (dbis.rwth-aachen.de/cms/projects/mi-mappa), das geeignete Akteure für komplexe Innovationen in der Medizin identifizieren sollte, unter PD Dr. Christoph Quix.
- **Benutzte Technologien:** Medical Text Analysis, Information Extraction, Text Mining, High-Performance Computing

- **Knowledge Mining und Assessment Group, TU Darmstadt** Darmstadt, Deutschland
Studentische Hilfskraft Mai 2015 - Jul 2015
- **Zusammenfassung:** Berechnung einer Vertrauensmaßnahme für die Zuverlässigkeit von Daten über Crowdsourcing-Enzyklopädie und kommerzielle Enzyklopädie unter Prof. Dr. Ulf Brefeld.
- **Benutzte Technologien:** Python, MySQL, Information Extraction, Text Mining, Conditional Random Fields, Wikipedia API

- **University of Wisconsin** Milwaukee, USA
Praktikant May 2014 - Jul 2014
- **Zusammenfassung:** Unter Prof. Hemant Jain wurden aktuelle Algorithmen zur Anonymisierung personenbezogener Daten in medizinischen Gesundheitsdaten analysiert und optimiert.
- **Bereiche der Entwicklung:** Java, HIPAA, Text Mining, Mu-Argus, MIST

Ausgewählte Publikationen

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.

Sprachkenntnisse

1. **Hindi:** Muttersprache
2. **English:** Fließend, TOEFL iBT Note: 118/120

3. **Deutsch:** *Verhandlungssicher*, Telc Deutsch als Fremdsprache B2, Note: 2,0

Referenzen

1. **Patrick Bornikoel, Data Reply GmbH**
p.bornikoel@reply.de
2. **PD Dr. Christoph Quix, Informatik 5 der RWTH Aachen**
quix@dbis.rwth-aachen.de, Telefon: +49 241 80 215 11
3. **Christoph Kreibich, Audi AG**
christoph.kreibich@audi.de