Nisshincho 23-01, #911
Kawasaki City, Tokyo, Japan

AVINASH MISHRA

HTTPS://AVINASH-MISHRA.GITHUB.IO

+81-70-2188-8241
avinash.mishra.2388@gmail.com

EMPLOYMENT

Data Scientist/Executive assistant to CTO — **paradigmshift.io, Japan** — **September 2016~**

Repchecker

- Repchecker is a B2B cloud based reputation checker of modern hotel industry.
- I am responsible for designing the search component of it. Customer can search and monitor their competitors based on time bound search. I am working on it from scratch, designed the architecture which involves technologies like Java, Cassandra multi-node cluster, Solr, Hadoop, Nginx and Jenkins.

Valuechain

- Valuechain is an NLP tool to process review data from OTAs.
- I am leading this project and we are team of 3 members.
- I have redesigned the architecture and we perform a lot of NLP techniques like Noise removal, Lexicon Normalization, Object standardization, Entity Parsing, Statistical featuring and text classification and matching etc.
- We recently started doing decision tree based supervised machine learning to predict customer's revisit.

Distributed web crawler / Data Lake

- I am also leading this project. We are a team of 5 engineers for it.
- I designed distributed web crawling architecture. It has centralize controller, distributed spiders run and distributed storage in Cassandra.
- Open source components usage : scrapy-redis, Scrapy-cluster, and Portia
- Data cleaner is another small module which we use to fetch data from Data Lake.

Data Migration / query optimization

- We use MariaDB for storage but system performance was quite slow. So I segregated unused and analytics related data into Cassandra and rest is in MariaDB.
- Time to time I also do query optimization and data modeling.

Software engineer (Big data department) — **Rakuten, Japan** — **Oct. 2014 – Aug. 2016**

GSP

- Global search platform was being used for almost all of the Rakuten services.
- Created automated testing framework (Ngauto) for distributed search services involving systems like Solr, Zookeeper, Cassandra and Hadoop.
- Created search peripheral components like dictionary compiler and word extractor to support search sub-functions like "did you mean", "related-words", "spell-check" and "auto-completion".
- DevOps : In-charge of CI/CD . Used docker, chef , Jenkins and some shell scripts for OS-provisioning and delivery pipeline.

Survey Panda

- Survey Panda is designed for PC support help desk feedback. The application is currently used by Rakuten employees at PC support help desk.
- Team size : 14
- I was part of backend team working on Spring boot (J2EE). My main task was to design survey database and validate survey form and post feedback form.

Dynamic Search UI

- Initial team member responsible for improving dynamic search UI of Rakuten ichiba.

Systems engineer **Infosys Ltd., India** **Mar. 2011 – Aug. 2012**

Molina Healthcare management

- Molina was a client of infosys and they wanted to upgrade their healthcare product. It was a pretty big project and we were team of 20+ engineers. I was responsible to manage users' health related information in database. I designed the DB in SQL server and performing stored procedures, triggers and functions to optimize it.
- Technologies : Java,Python, SQL server.

Employee SWAP portal

- Infosys is topmost company in India with more than 150,000 employee.
- India with big demographic region sometimes people gets posting at a place where they don't want to live. So I have designed one web based swap portal for employees who are interested to swap their posting locations.
- Technologies : Java, Jsp, Servlet.
- Team size : 5

Teaching assistant **VJTI Mumbai, India** **Sept. 2013 - June 2014**

- Courses: Data structures and algorithms, Software engineering, Mathematical foundation for computer science.

EDUCATION

Master of Technology **Software engineering, Department of CS (VJTI)** **Sept. 2012 – June 2014**

Bachelor of engineering **Computer science & engineering (RGPV)** **July 2006 – June 2010**

- *Post Graduate Coursework:* Human computer interaction; Design patterns; Artificial Intelligence; Computational Theory; Network analysis, Advanced Data Mining;
- *Post graduation research work* : Hadoop file system optimization.
- *Undergraduate Coursework*: Operating Systems; Databases; Algorithms; Digital electronics, Programming Languages; Comp. Architecture; Engineering Mathematics, Data Mining.

TECHNICAL EXPERIENCE

Projects

- **Content based image retrieval system** (2013). Used Java to implement exact match algorithm for comparing images and based on Euclidean distance retrieved optimal match.
- **Airline Analysis (POC) :** This project had following functionalities:
  It lists Airports operating in India, lists airline having zero stops, lists of all airlines operating, which country has highest Airports, list of active airlines in US.
- **e-Commerce site(POC) :** Build an e-Commerce shopping site to search products and display. Search text box supports auto-complete feature. Users can type minimum 3 characters to search for products that they want. Solr has been configured for auto suggester.
- **Youtube Data analysis(POC) :** Get top 5 categories with maximum number of videos uploaded, Get top 10 videos and most viewed videos.

Additional Experience and Awards

- **Rakuten Project Award** for GSP.
- HackerRank : Among top java developers in Japan. avinash_mishra02 **Current rank is 1**
- Stackoverflow profile : Avi
- Event organizer in Rakuten Technology conference 2014.
- Placement coordinator of software engineering branch during 2013-2014

Languages and Technologies

- Language : C , Java , Python, SQL
- Server : Apache tomcat, Jetty, Nginx
- Big data technologies : HDFS,Yarn, MapReduce,Spark, Kafka, Zookeeper
- Search technologies: Apache lucene, Solr
- Database : Mysql, Cassandra, MongoDB, MariaDB, SQL server, Redis
- DevOps/CI : Jenkins, chef, docker, Vagrant.
- IDE/Editors: Intellij Idea , PyCharm, Atom, Vim
- Tools : github, Jira, Confluence, Bitbucket.
- OS : Ubuntu, CentOs
- Cloud Platform: AWS, Digitalocean, Amazon Lightsail