Sunjit Rana

Senior Development Engineer (Exp: 5.6Years)

E-mail : sunjit92@gmail.com
Mobile : +91-8796862296

Summary:

- Senior data engineer with more than Five years of hands on experience in NLP and Data extraction through various source of information.
- Basic knowledge of Chatbot and Deep learning.
- Cohesive team worker, having strong analytical, problem solving and interpersonal skills.

Technical Skills:

Key Skills : Data Extraction, Natural Language Processing, Named Entity Recognition, Machine Learning
Technologies : Java, Python, R
Database : MySQL, Postgres
IDE and Tools : Eclipse, R Studio, WEKA, Stanford Parser
Packages Used : NLTK, Pandas, NumPy, Scikit learn, Distsance, Levenshtein, PyPDF2, PDFMiner.six PyTesseract, Tesseract ocr, Apache Tika, PDF2Image
Expertise : Ignio

Projects:

Pramati Technology, Hyderabad

Extracting information from E-mails (Jun 2018 – Present)

E-mail conversations about insurance policy discussion contains variety of information like Industry type, Insured details, Period of Insurance, Premium limit, Total premium, etc which will help to compute insurance premium when new customer comes from the same domain. This project is related to extract such useful information from previous interaction with other customers. Source is mail chain conversations with customer which is in the form of text written by human. Goal of the project is to help organization to increase conversion ratio from proposal to policy by analyzing previous data.

My Role:

I am working as senior data extractor in the team and my role is to use NLP techniques to extract information from mail conversations and to train algorithm in such a way that whenever same information comes in different order or with different tag, algorithm is able to identify that and extract information with high accuracy.

Extraction of Information from Invoice(Aug 2018- Present)

While processing the medical claims, insurance provider companies need to verify the claim amount which is claimed with the invoice attached. Currently this work is done manually, which result in

delay in processing of claim. Here we are trying to extract information from this invoices which are in image/ pdf format, and once extracted we are verifying the data against the claim raised.

- Libraries Used to get text from PDF: PyPDF2, PDFMiner.six, PyTesseract, Tesseract ocr
- Library to convert PDF file to Image: PDF2Image
- Algorithm used to correct words identified wrongly through Tesseract ocr: Affinity Propagation, Levenshtein ratio
- Language: Python

Cluster PDF Documents based on underlying structure(Jun 2018- Present)

Project is to create an application that can help to understand the structure of the documents, from large amount of PDFs with different layout is provided. This can help to reduce the initial time for screening of documents, to get good understanding of the structures lies within set of documents.

- Algorithms Used: K-means clustering, DB Scan, Hierarchical Clustering
- Libraries used: Pandas, NumPy, Scikit learn, NLTK

Tata Consultancy Services, Pune

Ignio for Batch (Sep 2017- May 2018)

Batch jobs and automated workloads are the backbone of many enterprise businesses. The complexity and sheer size of the workloads required for day-to-day operations often numbering in the hundreds of thousands—catches many operations teams by surprise. Ignio for Batch learns your business and keeps your processes running smoothly, saving time so you can focus on what's important: growing your business, not simply maintaining it.

My Role:

I had worked on end to end product deployment in client premises, and performing the analytics on client data. Giving the analysis results to them which will help them to improve the stability, Improves business agility by reducing SLA violations and predicting future batch job impacts with over 90% accuracy.

Extraction of Drugs from medical case reports (Oct 2016- Aug 2017)

Medical case reports contain detailed information about patients, their illnesses, and the medical procedures they undergo. Such information can provide useful inputs for the study of pharmacovigilance. Because of the plain text nature in which case reports are presented; coded summarization of case reports is often complicated and demand a niche NLP approach.

My Contribution:

- Worked on data cleaning and pre-processing
- Identify features to train SVM model and CRF for labeling
- Algorithms used: SVM, CRF, Association Rule Mining, Naive Bayes

Advancing AI to match human world class experts (Oct 2015 - Sep 2016)

While a computer matching human expertise has been a primary goal of AI research, it continues to remain elusive despite the enormous advances made in the field. Mind games have been viewed as a serious vehicle to promote and demonstrate AI research. We pick the card game of Bridge as our vehicle for exploring new ideas.

Mining process models from test cases (Dec 2013 – Sep 2015)

We had worked on mining software system model from set of available Test cases. Motivation behind our work is to comprehend the system from given set of Test cases. The mined model can help to gain understanding of the IT system from an outside in perspective as well as the manually

My Contribution:

I had worked in tool development. I have been involved in this project since it started and have written test scripts without having to execute them manually. worked through all the phases of the project. I had used supervised learning algorithm for classification, literature survey on information retrieval techniques etc. as a part of my work. I have also developed the GUI (Swing) & demonstrated to the client.

Internship Project:

Test Suite partitioning and scheduling to achieve parallelization

- Industrial Training in TRDDC, Pune(4 months)
- This project is related to software testing
- The main goal of this project is to execute test case parallel such a way that minimize the testing time of software

Certification:

- Oracle Certified Professional, Java SE6 Programmer with 98%
  Oracle, License 1Z0-851, July 2015

Publications:

1. Vipul Shah, Chetan Khadke and Sunjit Rana "**Mining process models and architectural components from test cases**" in the Software Testing, Verification and Validation Workshops (ICSTW), 2015 at Graz, Austria.
2. Chetan Khadke, Sunjit Rana and Vipul Shah "**Recovering Workflows from Functional Tests**" in the 23rd IEEE International Conference on Program Comprehension (ICPC), 2015 at Florence, Italy.
3. Vipul Shah, Sunjit Rana and Chetan Khadke "**Knowledge Transition: Discovering models from functional tests**" in the 26th IEEE International Symposium on Software Reliability Engineering (ISSRE), 2015 at Gaithersburg, USA.
4. Chetan Khadke, Sunjit Rana and Vipul Shah **"KnowX: Knowledge Acquisition from Functional Tests"** in International conference on Evaluation of Novel Approaches to Software Engineering (ENASE), 2016 at Rome, Italy.

5. Rajdeep Sarkar, Sunjit Rana, "**An approach to extract treatment/intervention details from medical case reports**" in proceedings of "9th International Conference and Exhibition on Pharmacovigilance & Drug Safety", which will held in July-2017, Munich, Germany.

- I have delivered invited talk at Fraunhofer Center for Experimental Software Engineering, Maryland on our work which I was presented at ISSRE – 2015.

Publication Link:

- http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Sunjit%20Rana.QT.&newsearch=true
- http://dl.acm.org/author_page.cfm?id=99658753446&CFID=755430290&CFTOKEN=16904169

Academic Record:

B.Tech. (CGPA: 8.5/10)
Computer Engineering Dharamsinh Desai University, Gujarat (2009-2013)

Personal Details:

Address: Sunjit Rana
2868, Ayodhya Nagar, Link road, Bharuch, Gujarat
Mail Id: sunjit92@gmail.com
Mobile: 8796862296/ 9033943145