KOUSTUV SAHA

341A, Tech Square Research Building
85 Fifth Street NW, Atlanta, GA 30308
(404) 692-9496
koustuv.saha@gatech.edu

SUMMARY

A Software Professional and Researcher in the field of Artificial Intelligence with primary interest in Natural Language Processing and Understanding, Machine Learning, Predictive Analytics, Information Retrieval and related domains. Acquainted with Big Data based tools and technologies. Pursues involvement with conceptualization and innovation to solve real-world problems

EDUCATION

Doctor of Philosophy Student, *Computer Science*   2016 -
Specialization: Social Computing, Social Dynamics and Well-being Lab
Advisor: Dr. Munmun De Choudhury
School of Interactive Computing, Georgia Institute of Technology

Bachelor of Technology (Hons.), *Computer Science & Engineering*   2008-12
B.Tech Project: Ontology based Semantic Information Retrival System
Guide: Dr. Sujoy Ghose
Computer Science & Engieneering, Indian Institute of Technology Kharagpur

PATENTS

- *Bose Arnab, Saha Koustuv, Das Sumit, Biswas Nirmalendu*, **A System and Method for Predicting Compliance Issues**,
  US Provisional Patent filed, February 2014

- *Patel Labhesh, Saha Koustuv, Das Sumit, Biswas Antarip*, **A System and Method for Sentiment Analysis of Text**,
  US Provisional Patent filed, August 2014 / US Patent filed, August 2015

- *Patel Labhesh, Saha Koustuv, Das Sumit, Biswas Antarip*, **A System and Method for scalable and continuous learning and annotation of data**,
  US Provisional Patent filed, March 2015

TECHNOLOGY SKILLS

- Proficient in software development using Java. Familiar with C, C++, R, Python, Neo4j, MySQL, MongoDB, Hadoop and MapReduce

- Familiar with NLP Techniques like POS, NER, linguistic and grammatical rules, semantic role labelling, tree and dependency parsing, Word-Vectors, WordNet, DMOZ, Lexicons and Ontology, domain modeling etc

- Experienced with Machine Learning and predictive modeling algorithms and methodologies

EXPERIENCE

Senior Researcher   December 2012 - August 2016
Abzooba Inc   Abzooba X

- Worked as the Product Development Lead of Abzooba's flagship product, Xpresso, which does intelligent analysis and semantic understanding for a multitude of natural language problems, like aspect based expressions

classification (sentiment, statement type, emotion), resume scoring, document classification, social media trend detection and intent classification
- Worked as the Product Development Lead of Synaptica, a natural language based search engine for ecommerce platforms
- Worked on sub-problems of text processing, hashtag processing, LSH-based vector similarity, Feature-Opinion based parser, ontology augmentation heuristics, ngram usage for Candidate lexicon generation and Label Sequence based sequence learning
- Was instrumental in the conceptualization and implementation of Continuous Learning System (CLS)
- Worked with latest NLP techniques like Semantic Role Labelling, Dependency Parsing, linguistic rules and feature engineering for Machine Learning Models for expressions classifier
- Used word-vectors to address unsupervised natural language problems like aspect categorization, semantic information retrieval and document classification
- Achieved highest rank of 23 (700 terams) in Kaggle's global sentiment analysis problem
- Played a key role in the innovation and implementation of TagIT (a human annotator tool), ontology builder (semi-automated domain modelling) and QAtools (automated testing)
- Was involved with development, customizations and delivery of solutions on top of Xpresso to various clients
- Played a key part in the innovation and implementation of Compliance Risk IndeX (CRIX) - a patented solution for predicting Retail Health Index, using mixed statistical advanced analytics modeling, leveraging structured, unstructured and real-world data. This solution was ranked 1st at Hackathon by a US retail giant
- Involved briefly on Large Group Predictive Modeling to predict memberwise monthly costs of Health Insurance
- Innovated DeaDRIX (Dealer Distributor Relative Index) for relative ranking of Dealer wrt Distributor based on sales statistics
- Visiting faculty member at Praxis Business School and IMS Institute for courses on Big Data, Retail Analytics and NLP
- Recognized and awarded for distinctive contribution towards analytics solutions

Junior Research Associate   August 2012 - November 2012
Infosys Ltd   Infosys Labs

- Conceptualized heuristics and developed Java based tool for information extraction from unstructured financial analysis data
- Contributed in preparing input dataset and ideating on a semantic similarity based approach to recommend trending projects

Research Intern   May 2014 - February 2016
Siemens Information Systems Ltd   Corporate Technologies

- Set up a Hadoop cluster of 30 computers, installed Mahout and Katta for running jobs
- Developed a Java based application to access Lucene indexes to search and analyze their content based on parameters
- Involved with research on algorithm and applications of Map-Reduce. Delivered a talk on Indexing in Information Retrieval: Tools and Techniques