


Nausheen Fatma

MS BY RESEARCH STUDENT, COMPUTER SCIENCE, IIIT-HYDERABAD


EDUCATION **International Institute of Information Technology, Hyderabad**
MS By Research, Computer Science and Engineering *Aug' 14 - May' 17 (Expected)*
CGPA: 8.5/10

Jamia Millia Islamia (Central University), New Delhi
B.Tech, Computer Engineering *Aug' 08 - Jun' 12*
CGPA: 8.54/10 (First Divison with Honors)

RESEARCH INTERESTS Natural Language Processing and Artificial Intelligence

PUBLICATIONS **Nausheen Fatma, Manoj K. Chinnakotla, Manish Shrivastava**
The Unusual Suspects: Deep Learning Based Mining of Interesting Entity Trivia from Knowledge Graphs. In: Proceedings of the the Thirty-First AAAI Conference, AAAI-17. 

Nausheen Fatma, Manoj K. Chinnakotla, Manish Shrivastava
Relevance Scoring of Triples using Ordinal Logistic Classification. To appear In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM Cup 2017*

Nausheen Fatma, A Kamineni, Arpita Das, Manoj K. Chinnakotla, Manish Shrivastava
IIITH at BioASQ Challenge 2015 Task 3a: Extreme Classification of PubMed Articles using MeSH Labels. In: *Conference and Labs of the Evaluation Forum, CLEF 2015.* 

AWARDS & ACHIEVEMENTS - Travel Grants from **Google, Microsoft** and **ACM** to present at *AAAI-17*.
- **100 % Tuition Fee Scholarship** for pursuing Masters degree by NLP-MT Lab, IIIT-H.
- 94.17 percentile in **GATE 2014**.

PROJECTS **Trivia Mining from Knowledge Graphs (Thesis project):** *Mar '16 - Present*

The task was to mine interesting/trivia facts from knowledge graphs such as DBpedia. Used Deep Learning and Machine Learning approaches like CNN, SVM, Gradient Boosting Trees and achieved best F1 score of 0.81 and 0.65 for two domains: "Bollywood Actors" and "Music Artists". This work was published in AAAI-17.

Relevance Scoring of Triples using Ordinal Logistic Classification: *Nov '16 - Jan '17*

The task was to compute relevance scores or rank for triples from a knowledge graph. The triples consisted of type-like relations for two categories: "Profession" and "Location". Used Ordinal Logistic classification model and achieved an overall accuracy score of 0.73 and Kendall's tau score of 0.36. This work is accepted for publication in WSDM Cup 2017.

Wikipedia Search Engine from Scratch: *Jan '16 - Mar' 16*

This was a course project with consisted of two phases-Indexing and Retrieval. Using a 46 GB Wikipedia dump, I implemented a fully functional search engine and used Okapi BM25 for ranking which gave top 10 results for an input query in 0.5 second. I employed Map-Reduce framework for speeding the Indexing task.

Question classification using Machine Learning :

Aug '15 - Sep '15

As a part of course project, we implemented the research paper “*Learning Question Classifiers*” by Xan Li and Dan Roth. We used Linear SVM classifier on the TREC dataset and got the accuracy of 91% using features Bag of Words, POS, NER, and Chunk features.

Multi-label classification for PubMed articles (Team of 3):

Nov '14 -Feb '15

The task was to do multi-label classification and identify the MeSH labels for PubMed documents. We used three approaches K-Nearest Neighbours, IDF Ratio approach, and Extreme classification using FastXML. This work got accepted and published in the CLEF 2015.

Study of Convolution Neural Networks for basic NLP tasks:

Nov '15 - Dec '15

POS tagging of 5 Indian languages (Hindi, Urdu, Marathi, Malayalam, Bangla) using the SENNA tool and study the working of CNN for basic NLP tasks as described in the research paper “*Natural Language Processing (almost) from Scratch*” by Robert Collobert.

Bag of Words Meets Bags of Popcorn (Kaggle project) :

Oct '16 - Nov '16

Sentiment analysis of IMDB movie reviews using Machine Learning with various feature representation such as Bag of Words and Word embeddings to predict the polarity/ sentiment of user reviews.

Semantic Job recommendation system (Team of 3):

Mar '16 - Apr '16

We developed a system where given a job description, the system would return a ranked list of CVs suitable for that job position. Due to limited dataset, we implemented a simple TF-IDF based system.

For more project listings and code, visit my github page [here](#).

**WORK
EXPERIENCE****National Informatics Centre, New Delhi***Java Programmer*

Sep' 12 - Jul' 14

Developed major core functionalities: DashBoard (Home page), Audit Trail, Alerts & Notification, Integrated Holiday Calendar, Reports for the e-Leave (Leave Management System) module of e-Office project. For more details about the project, see: <https://eoffice.gov.in/>.

International Institute of Information Technology, Hyderabad*Teaching Assistant (Natural Language Processing)*

Aug' 15 - Dec' 15

With over 75 students, I was mentoring the students during the course and played the lead role in evaluation of assignments and exams.

NLP-MT Lab, International Institute of Information Technology, Hyderabad*Research Assistant (under Prof. Dipti Misra Sharma)*

Aug' 15 - Current

Support tasks like: Implementing module for next word prediction and spell checker using Bigram Language Modelling, debugging and testing the output of *Hindi-Urdu ILMT project*, handling student registrations and fees payments at *Panini Linguistics Olympiad*, etc.

SKILLS

Python, Scikit-learn, Keras, SPARQL, Flask, Java, Hadoop