

# Ketan Kumar Todi

<https://www.linkedin.com/in/ketan-kumar-todi/>

<https://todiketan.github.io>

Email : ktodi@andrew.cmu.edu

Mobile : +1-412-616-6134

Google Scholar

## EDUCATION

### Carnegie Mellon University, School of Computer Science

Pittsburgh, PA

*Master of Science in Intelligent Information Systems; GPA: 4.08/4.00*

*Dec 2022*

Courses: Advanced Natural Language Processing, Intro to Machine Learning, Search Engines

### Manipal Institute of Technology

Manipal, India

*Bachelor of Technology in Computer Science and Engineering; GPA: 9.36/10.0 (Top 5%)*

*Jul 2019*

## EXPERIENCE

### Carnegie Mellon University

Pittsburgh, PA

*Graduate Research Assistant - Prof. Graham Neubig*

*Aug 2021 – Present*

- Exploring **paragraph pruning** methods using attention scores from retrieval model to improve efficiency and make the model scalable for multiple hops. Achieved 80% F1 score of the original QA model by using 50% of passage tokens.
- Improving multi-hop question answering performance by enhancing second hop retrieval process through **explicit query reformulation**.

### Taiger

Singapore

*Research Engineer*

*Jun 2019 – July 2021*

- Developed an active-learning based trainable NER, relation extraction and relation classification model using BERT for **semi-automatic ontology generation**, cutting down data requirements by **upto 50%** for ontology generation from domain specific documents; awarded as the best performer in the R&D team of 20 employees for this project.
- Designed a CRF based stacking ensemble model for **named entity extraction** from legal documents, where ensemble model had a **3% gain** over the best performing individual model.
- Developed a fast-text based **multilabel text classification** system. Worked on **few-shot** intent classification for Arabic and Chinese using kNN and Universal Sentence Encoder achieving **70% F1 score**, 10% gain over baseline model.

### International Institute of Information Technology, Hyderabad (IIIT-H)

Hyderabad, India

*Research Intern - Prof. Manish Shrivastava*

*May 2018 – June 2018*

- Devised a cross-sentence attention based neural network for **Semantic Textual Similarity** task, achieving an accuracy of **0.76**, close to the state-of-the-art score of **0.81** at the time of internship.

*Research Intern - Prof. Dipti Misra Sharma*

*Dec 2017 – Jan 2018*

- Developed part of speech tagging (POS) module for an Indian Vernacular language, named Kannada using machine learning (**mainly CRF and SVM**) and **bi-LSTM**, achieving state-of-the-art results of 92% F1 score, which was 6% higher as compared to the previous results.

## PROJECTS

### Template Based NER using BART - Advanced NLP

Ongoing

- Proposed a **prompt** based sequence-to-sequence NER model using BART. The decoder prediction followed **fixed virtual tokens template** learnt through back-propagation.

### Search Engine

Fall 2021

- Developed a minimal search engine supporting exact and partial match (**BM25 and Indri**), **LeToR**, query reformulation using pseudo relevance feedback (**PRF**) and **PM2 and xQuad** diversity algorithms using Java.

## PUBLICATIONS

- **Knowledge Graph Generation with Deep Active Learning:** Abhishek Pradhan\*, Ketan Kumar Todi\*, Anbarasan Selvarasu, Atish Sanyal. IJCNN, Glasgow, UK, 2020
- **Decentralizing Air Traffic Flow Management with Blockchain-based Reinforcement Learning:** Ta Duong, Ketan Kumar Todi, Umang Chaudhary, Hong-Linh Truong. INDIN, Finland, 2019
- **Building a Kannada POS Tagger Using Machine Learning and Neural Network Models:** Ketan Kumar Todi\*, Pruthwik Mishra\*, Dipti Misra Sharma. CICLING, Vietnam, 2018

## PROGRAMMING SKILLS

- **Languages:** : C++, Python, Java, LATEX
- **ML Libraries:** : Pytorch, Tensorflow, Spacy, Pandas, NLTK, scikit-learn