

# TANVI SAHAY

BRITTANY MANOR DRIVE, AMHERST, MA - 01002

[tsahay@cs.umass.edu](mailto:tsahay@cs.umass.edu) | [G](#) | [in](#) | [webpage](#)

## EDUCATION

MAY 2018	University of Massachusetts Amherst, MA, Master of Science, Computer Science	4.0
JUNE 2016	Birla Institute of Technology, India, Bachelor of Engineering	3.0

## SKILLS

Python, Tensorflow, Javascript, D3, Numpy, scikit-learn, nltk, Stanford CoreNLP, SQL, Django, MongoDB, Git, LaTeX

## WORK EXPERIENCE

- Sep 2017 - Present **Graduate Research | Information Extraction and Synthesis Lab | UMass Amherst**
- Explore deep learning based models for obtaining text representation helpful in modeling the expertise of researchers for matching them to papers they are qualified to review.
  - Study and Develop deep learning models for the purpose of key phrase extraction to assist in better expertise modeling.
- Feb 2017 - Aug 2017 **R&D Intern and Independent Study | Lexalytics Inc.**
- Successfully implemented several NLP baseline as well as word2vec and deep learning based models to obtain fixed-dimensional distributed representations for phrases for the purpose of clustering phrases based on relatedness
  - Experimented with KMeans, DBSCAN, Heirarchical and Spectral clustering and several cluster evaluation techniques for obtaining phrase clusters coherent to human evaluators.

## SELECTED RESEARCH PROJECTS

- Sep 2017 - Dec 2017 **Automatic Colorization of Videos**
- Compared the performance of existing image colorization architectures for automatically colorizing video frames using pixel wise rmse and colorization quality measured by human evaluators
  - Experimented with an RCNN architecture to account for consistency between consecutive frames and used a re-weighted class-rebalancing loss to avoid desaturated colorization of grayscale images
- Feb 2017 - May 2017 **Schema Matching using Machine Learning**
- Engineered custom features to represent schema names and employed Self Organizing Maps and Gaussian Clustering to cluster similar schema names
  - Performed within-cluster one-to-one matching using edit distance and introduced the idea of domain-based global dictionary for the purpose of one-to-many schema matching
- Oct 2016 - Dec 2016 **Sentence Generation using Fan Theories**
- Employed CoreNLP package for tokenization, relation extraction, PoS tagging and Named Entity Recognition on a database of Fan Theories of Game of Thrones
  - Performed noise removal using OpenIE and used Bigram, HMM and character-level LSTMs for sentence generation.
  - Analyzed results based on overall coherence, general fluency and information content with and without considering domain knowledge.
- Oct 2016 - Dec 2016 **Architecture Classification for Indian Monuments using ORB features**
- Extracted ORB features of monument images and performed architecture classification using KNN, Logistic Regression, SVM and Random Forests.
  - Compared image-wise classification using different supervised techniques with descriptor-wise classification using KNNs.

## SELECTED PUBLICATIONS

- A. Aggarwal, T. Sahay, A. Bansal and M. Chandra, "Grid search analysis of nu-SVC for text-dependent speaker-identification," 2015 Annual IEEE India Conference (INDICON), New Delhi, 2015. *Best Paper Award*
- T. Sahay, A. Aggarwal, A. Bansal and M. Chandra, "SVM and ANN: A comparative evaluation," 2015 International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2015.

## OTHER EXPERIENCE

- Grader for the graduate level CS 589 Machine Learning course for Spring 2016 and Fall 2017
- Poster presentation at the Women in Data Science Conference, Boston, January 2018.