

# 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, PyTorch, Javascript, D3, Numpy, scikit-learn, nltk, Stanford CoreNLP, SQL, Django, MongoDB, Git, LaTeX

## WORK EXPERIENCE

---

Oct 2018 -Present	<b>AI Researcher   AlphaSense Oy</b> <ul style="list-style-type: none"><li>Research and implement Artificial Intelligence, Natural Language Processing and Machine Learning solutions</li></ul>
Feb 2018 -May 2018	<b>AI Research Intern   AlphaSense Inc.</b> <ul style="list-style-type: none"><li>Worked in collaboration with the Innovation Team to develop automated Deep Learning and NLP based solutions for various company projects.</li></ul>
Sep 2017 -Feb 2018	<b>Graduate Research   Information Extraction and Synthesis Lab   UMass Amherst</b> <ul style="list-style-type: none"><li>Explored machine learning techniques for obtaining interpretable models of researcher expertise to match new scientific works with potential reviewers.</li><li>Explored deep learning models for the purpose of key word/phrase extraction from research papers, to assist in better expertise modeling.</li></ul>
Feb 2017 -Aug 2017	<b>R&amp;D Intern and Independent Study   Lexalytics Inc.</b> <ul style="list-style-type: none"><li>Successfully implemented several NLP baseline and word2vec and deep learning based models to obtain fixed-dimensional distributed representations of phrases for the purpose of clustering semantically-related phrases</li><li>Experimented with KMeans, DBSCAN, Hierarchical and Spectral clustering and several cluster evaluation techniques for obtaining phrase clusters coherent to human evaluators.</li></ul>

## SELECTED RESEARCH PROJECTS

---

Sep 2017 -Dec 2017	<b>Automatic Colorization of Videos</b> <ul style="list-style-type: none"><li>Compared the performance of existing image colorization architectures for automatically colorizing video frames using pixel wise rmse and colorization quality measured by human evaluators</li><li>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</li></ul>
Feb 2017 -May 2017	<b>Schema Matching using Machine Learning</b> <ul style="list-style-type: none"><li>Engineered custom features to represent schema names and employed Self Organizing Maps and Gaussian Clustering to cluster similar schema names</li><li>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</li></ul>
Oct 2016 -Dec 2016	<b>Sentence Generation using Fan Theories</b> <ul style="list-style-type: none"><li>Employed CoreNLP package for tokenization, relation extraction, PoS tagging and Named Entity Recognition on a database of Fan Theories of Game of Thrones</li><li>Performed noise removal using OpenIE and used Bigram language model, Hidden Markov Model and character-level LSTM for sentence generation.</li><li>Analyzed results based on overall coherence, general fluency and information content with and without considering domain knowledge.</li></ul>
Oct 2016 -Dec 2016	<b>Architecture Classification for Indian Monuments using ORB features</b> <ul style="list-style-type: none"><li>Extracted ORB features of monument images and performed architecture classification using KNN, Logistic Regression, SVM and Random Forests.</li><li>Compared image-wise classification using different supervised techniques with descriptor-wise classification using KNNs.</li></ul>

## 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.

## ADDITIONAL EXPERIENCE

- 
- Grader for the graduate level CS 589 Machine Learning course for Spring 2017, Fall 2017 and Spring 2018