

Zaid Sheikh

Last updated: 22 Sept 2021

sheikh.zaid[at]gmail.com
linkedin.com/in/zaidsheikh

WORK EXPERIENCE

Carnegie Mellon University, Pittsburgh Jan 2016 - Present

Senior Research Programmer

- (Mar 2021 - Present) DSTA (Defence Science and Technology Agency) Singapore project on social media summarization and narrative generation for humanitarian aid and disaster response
- (Jan 2019 - Present) DARPA AIDA project: Develop and maintain software for building a probabilistic knowledge base composed of entities, events and relationships between them via automated analysis of text and multimedia data
- (Sep 2019 - Mar 2021) DARPA KAIROS project: Develop software that identifies, links and temporally sequences sub-events to understand complex events in text and multimedia data
- (Jan 2016 - Sep 2019) DARPA LORELEI project: Design software architectures for text analytics to identify topics, names, events, sentiment and other elements of information in low-resource foreign language sources

Voci Technologies, Pittsburgh

Aug 2014 - Dec 2015

Speech Engineer

- Built a speech analytics and data visualization application (V-Spark Voice Analytics™) using state-of-the-art speech recognition and text analysis technologies
- Devised and put into effect a semi-automated procedure to correct transcriptions of audio files in training and testing datasets
- Also worked on speaker diarization, improving automatic transcription of number sequences in conversational speech, music and noise detection etc.

Carnegie Mellon University, Pittsburgh

Sept 2013 - July 2014

Research Programmer

- Built low-resource and zero-resource Spoken Term Detection systems for multiple languages as part of the IARPA Babel Program
- Put together a prototype system integrating all of our work on the Babel project and delivered it to IARPA at the end of the program

Carnegie Mellon University, Pittsburgh

Sept 2011 - Aug 2013

Graduate Research Assistant (part-time)

- Worked on improving acoustic modeling (Hidden Markov Models, Deep Neural Networks) for conversational speech recognition
- Built low-resource and zero-resource Spoken Term Detection systems for multiple languages as part of the IARPA Babel Program

Agnity India Pvt Ltd, Noida

June 2010 - July 2011

Software Engineer

- Developed SIP VoIP clients for Blackberry and Android smart-phones
- Worked on various Java EE and GWT (Google Web Toolkit) based projects

National Institute of Technology, Allahabad

June 2007 - May 2010

Full Stack Web Developer (part-time)

- Development/maintenance of the Student Admission system and the Course Management system using the LAMP (Linux, Apache, MySQL, PHP) stack.

EDUCATION	Carnegie Mellon University, Pittsburgh, PA M.S. in Computer Science - Language Technologies	Aug 2013 3.50/4.33
	National Institute of Technology, Allahabad, India B.Tech. in Computer Science	May 2010 9.23/10.00
SKILLS	Proficient in Python, JavaScript, Linux Shell Scripting, Java. Familiar with NLTK, Numpy, Scipy, Scikit-learn, PyTorch. Web development: JavaScript, Node.js, jQuery, SQL, PHP, HTML/CSS, Bootstrap.	
PUBLISHED PAPERS	Reducing Confusion in Active Learning for Part-Of-Speech Tagging - A. Chaudhary, A. Anastasopoulos, Z. Sheikh, G. Neubig - <i>TACL 2021 (To Appear)</i>	
	Automatic Extraction of Rules Governing Morphological Agreement - A. Chaudhary, A. Anastasopoulos, A. Pratapa, D. Mortensen, Z. Sheikh, Y. Tsvetkov, G. Neubig - <i>EMNLP 2020</i>	
	A Little Annotation does a Lot of Good: A Study in Bootstrapping Low-resource Named Entity Recognizers - A. Chaudhary, J. Xie, Z. Sheikh, G. Neubig, J. Carbonell - <i>ACL 2019</i>	
	OPERA: Operations-oriented Probabilistic Extraction, Reasoning, and Analysis - E. Hovy, J. Carbonell, H. Chalupsky, A. Gershman, A. Hauptmann, F. Metze, T. Mitamura, Z. Sheikh, A. Dangi, A. Chaudhary, X. Chen, X. Kong, B. Huang, S. Medina, H. Liu, X. Ma, M. Ryskina, R. Sanabria, V. Gangal - <i>TAC 2019</i>	
	The ARIEL-CMU Systems for LoReHLT18 - A. Chaudhary, S. Dalmia, J. Hu, X Li, A. Matthews, A. Obaja, N. Otani, S. Rijhwani, Z. Sheikh et al. - <i>arXiv:1902.08899 (2019)</i>	
	Parser combinators for Tigrinya and Oromo morphology - P. Littell, T. McCoy, N. Han, S. Rijhwani, Z. Sheikh, D. Mortensen, T. Mitamura, L. Levin - <i>LREC 2018</i>	
	The ARIEL-CMU Situation Frame detection pipeline for LoReHLT16: A model translation approach - Littell, P., Tian, T., Xu, R., Sheikh, Z., Mortensen, D., Levin, L., Tyers, F., Hayashi, H., Horwood, G., Sloto, S., Tagtow, E., Black, A., Yang, Y., Mitamura, T., and Hovy, E. - <i>Machine Translation (2018)</i>	
	Semi-Supervised Training in Low-Resource ASR and KWS - F. Metze, A. Gandhe, Y. Miao, Z. Sheikh et al. - <i>ICASSP 2015</i>	
	Multilingual Deep Bottle Neck Features - A Study on Language Selection and Training Techniques - M. Müller, S Stüker, Z. Sheikh, F. Metze, A. Waibel - <i>IWSLT 2014</i>	
	Models of tone for tonal and non-tonal languages - F. Metze, Z. Sheikh, A. Waibel, J. Gehring, K. Kilgour, Q. Nguyen, V. Nguyen - <i>ASRU 2013</i>	
	Identification and Modeling Of Word Fragments in Spontaneous Speech - Yulia Tsvetkov, Zaid Sheikh, Florian Metze - <i>ICASSP 2013</i>	
	A trigram part-of-speech tagger for the Apertium free/open-source machine translation platform - Zaid Sheikh, Felipe Sanchez-Martinez - <i>International Workshop on Free/Open-Source Rule-Based Machine Translation, Spain, 2009</i>	

REPORTS

Zaid Sheikh, Zhen-zhong Lan, Shourabh Rawat “Feature Learning for Multimedia Event Detection using Neural Networks” - *10-708 Probabilistic Graphical Models project report, Spring 2013*

Zaid Sheikh, Dandan Zheng “Social Tag Prediction for Stackoverflow Posts” - *10-802 Social Media Analysis project report, Fall 2012*

Yulia Tsvetkov, Zaid Sheikh “Improving Decision Trees for Acoustic Modeling” - *10-701 Machine Learning project report, Spring 2012*

Zaid Sheikh “Language Independent Query-by-Example Spoken Term Detection by Cross-Language Phonetic Decoding” - *11-751 Speech Recognition and Understanding project report, Fall 2011*

RELEVANT COURSES

Algorithms for NLP, Speech Recognition and Understanding, Machine Learning, Language and Statistics, Social Media Analysis, Design and Engineering of Intelligent Systems, Machine Learning with Large Datasets, Probabilistic Graphical Models

AWARDS

- Research Fellowship, Carnegie Mellon University, 2011-2013
- Google Summer of Code (summer of 2009 and 2010)
- Summer Research Fellowship, Indian Institute of Technology Bombay, 2009