

# Ankur Gandhe

ankurgan@cs.cmu.edu  
(412)-352-8522

<http://www.cs.cmu.edu/~ankurgan>

<b>Interests</b>	Language modeling, Information retrieval, Machine translation	
<b>Education</b>	Masters in Language Technologies, Carnegie Mellon University	July 2014 (Expected)
	B.Tech, Electronics and Communication Engineering, Indian Institute of Technology Guwahati	May 2010
<b>Technology Skills</b>	<u>Languages:</u> Java, Python, C++, VB .NET, VHDL, 8086 and TMS320C6416 Assembly <u>Software:</u> UIMA, LATEX ,MATLAB, HTK, CCStudio, HFSS,LTSpice, Visual Studio	
<b>Relevant Courses</b>	Machine Learning, Algorithms for NLP, Probabilistic Graphical Models, Information retrieval, Intermediate Statistics, Speech Recognition and Understanding, Deep Learning, Software Engineering, Speech processing, Information theory and coding, Pattern Recognition	
<b>Experience</b>	<b>Graduate Research Assistant, CMU</b>	Sept 2012 - Present
	<ul style="list-style-type: none"><li>• Working on context-aware Neural Network Language modeling (NNLM) for robust speech recognition, and optimizing them for key word search in the DARPA sponsored BABEL project. Network training is performed on a GPU using the theano library for python. Implemented the feed forward and recurrent NNLM in Janus recognition toolkit, and a context-aware NNLM in Hydra, a GPU-based decoder, for decoding and re-scoring.</li><li>• Working on automatically crawling, cleaning and filtering web text for language modeling of languages with low resources to improve keyword search.</li><li>• Built a question answering software for Machine Reading Task using the UIMA framework to answer multiple-choice questions following a medical passage. Various similarity metrics were implemented and then combined using UIMA to yield optimum performance on a development set.</li></ul>	
	<b>Software Engineer, IBM Research - India</b>	July 2010 - July 2012
	<ul style="list-style-type: none"><li>• Worked at IBM Research India as a software engineer in the Human Language Technologies group, to improve the machine translation system at IBM, with a focus on Indian languages. Worked on specific problems of handling word reordering and morphology in machine translation for Hindi-English systems.</li><li>• Worked at implementing an Italian to English translation assist tool as a plug-in to Microsoft Word for IBM GDLT group.</li><li>• Worked at building a system to efficiently extract answers from discussion forums and rank them according to relevance and correctness.</li></ul>	
	<b>Student Researcher, Telecom Bretagne, France</b>	May-July 2009
	<ul style="list-style-type: none"><li>• Applied FAME strategy to extract the AM and FM cues for spectral reduction of speech in speech processors and improve the performance of cochlear implants that use it. The algorithm was developed in C floating point and fixed point</li></ul>	

- Implemented the above algorithm real time on DSP TMS320C6416 and optimized the assembly code based on number of sub-bands and bit-rate for minimum processing time and minimum input-to-output delay

**Student Researcher**, IIT Delhi

May-July 2008

- Developed an automatic speech segmentation software using 5-state left to right Gaussian mixture models for English language phonemes using HTK toolkit and MATLAB for large vocabulary database. Classified the phonemes based upon voiced/unvoiced, stop/fricative, etc. and evaluated the mean error of 10ms by testing the software against hand-labeled data

## Publications

### **Optimization of Neural Network Language Models for Keyword Search**

To appear in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2014, Florence, Italy

### **Using Web Text to Improve Keyword Spotting in Speech**

Automatic Speech Recognition and Understanding (ASRU) 2013, Olomouc, Czech Republic

### **Hypothesis Refinement Using Agreement Constraints in Machine Translation**

International Joint Conference on NLP (IJCNLP) 2013 , Nagoya, Japan

### **Domain Adaptive Answer Extraction for Discussion boards**

World Wide Web Conference (WWW) 2012, Lyon, France

### **Clause-Based Reordering Constraints to Improve Statistical Machine Translation**

International Joint Conference on NLP (IJCNLP) 2011, Chaing Mai, Thailand

### **Handling verb phrase morphology in highly inflected Indian languages for Machine Translation**

International Joint Conference on NLP (IJCNLP) 2011, Chaing Mai

### **A Word Reordering Model for Improved Machine Translation**

Empirical Methods on Natural Language Processing (EMNLP) 2011 , Edinburgh

## Patents

Enhancing Posted Content in Discussion Forums

Filed in June 2012 by IBM

Providing an Uninterrupted Reading Experience

Filed in May 2012 by IBM

Conveying Context Sensitive Information for a Conversation

Filed in April 2012 by IBM

## Academic Honors

- Qualified for the Indian National Math Olympiad, INMO 2005
- Top 1% of the candidates in the Indian National Physics Olympiad, 2006
- Ranked 27 at the National Talent Search Examination (NTSE), 2003
- Represented Bangalore at Prof. Brahmo-Prakash Memorial Materials quiz, 2004,2005