CURRICULUM VITAE

ADDRESS: 407 S AIKEN AVE, APT 4,
PITTSBURGH, PA – 15232-1207, USA

MOHIT KUMAR

EMAIL: MOHITKUM@GMAIL.COM

RESEARCH INTERESTS

Summarization, Natural Language Processing, Machine Learning, Speech Technologies, Bio-informatics

PUBLICATIONS

- Masters Thesis, "Abstractive Summarization with 'Briefing' Focus" (In preparation).
- M. Kumar, N. Garera, A. I. Rudnicky, "Learning from the Report-writing Behavior of Individuals", accepted at International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, Jan'07.
- M. Kumar, N. Garera, A. I. Rudnicky, "A Briefing Tool that Learns Individual Report-writing Behavior", accepted at IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Washington, DC, Nov'06.
- D. Huggins-Daines, M. Kumar, A. Chan, M. Ravishankar and A. I. Rudnicky, "POCKETSPHINX: A Free, Real-time Continuous Speech Recognition System for Hand-held devices", ICASSP'06, Toulouse, France.
- M. Kumar, N. Rajput, A. A. Nanavati, P. Kankar and R. Dahiya "SAMVAAD: Speech Application Made Viable for Access-Anywhere Devices", IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMOB'05), Montreal, Canada.
- M. Kumar, N. Rajput and A. Verma, "A large vocabulary continuous speech recognition system for Hindi", *IBM Journal of Research and Development,* 2004.

PATENT

"A System and Method for Device-specific Dialog Call-flow Optimisation". Filed with USPTO – April, 2004.

EDUCATION

PhD Candidate — Aug'06-present — Carnegie Mellon University
Language Technologies Institute, SCS
Cumulative QPA: 3.86/4

Master of Science — Aug'04-present — Carnegie Mellon University
Language Technologies Institute, SCS
Cumulative QPA: 3.86/4 (Expected Graduation – Dec'06)

Bachelor of Technology (Hons.) — 1998-2002 — IIT Kharagpur
in *Computer Science and Engineering*
GPA: 8.32/10 (absolute grading)
Major GPA: 3.5/4

RESEARCH/WORK EXPERIENCE

1. **Graduate Research Assistant, Project RADAR, advisor: Dr. Alex I. Rudnicky, Carnegie Mellon University** – Aug' 04 – present

- I am currently working for my Masters thesis in the field of Summarization under the RADAR project. In particular, I am interested in the problem of 'abstractive' summarization which involves abstracting information from the given text, for e.g. a set of emails, in the form of a concise summary. I am also interested in utilizing multi-modal information for the same. We have currently built a preliminary system for 'extractive' summarization where the summary is presented as a set of sentences 'extracted' from the given text, which is a step towards abstractive summarization.

2. **Technical Staff Member, IBM India Research Lab (Delhi)** – Aug'03 – Aug'04

- I worked on building an Indian English Text-to-Speech system. More specifically, the research problem was to support Indian names that involve Hindi-specific phonemes.
- I explored the implication of speech technologies on pervasive devices. With the proliferation of pervasive devices, ubiquitous access through multiple modalities is becoming imperative. We addressed the problem of speech application adaptation by dialog call-flow reorganisation for pervasive devices with different memory constraints.

3. **Research Trainee, Speech Group, IBM India Research Lab (Delhi)**: July'02 – Aug'03

I worked on building a continuous speech, speaker-independent Hindi Speech Recognition (Desktop application) system.

- My main contribution had been to develop a novel technique of generating phonetic spellings of words to build the vocabulary in the system. I also designed and built the feature of adding words to the vocabulary software. The first version of the research product is now in the market. It has a vocabulary of 75K words and with speaker training, is achieving accuracy levels beyond 90%.
- I researched on improving the statistical (n-gram) language models using Semantic information.

OTHER & CLASS-RELATED PROJECTS DONE/DOING

At CMU:-

- PocketSphinx: I am working on porting SPHINX 2, CMU's speech recognition system, for WinCE operating system meant for embedded/handheld devices.
- Chemical Shift Prediction: I am working with Dr. Chris Langmead on predicting chemical shifts for proteins given their 3-d structure with the motivation that accurate chemical shift prediction may then be used in predicting protein structures and other related problems.
- Unsupervised Entity Association: In the project course, 'Read The Web' (Dr. Tom Mitchell), I am working with my team on extracting associations between named entities from MEDLINE corpora and the web using unsupervised metrics.
- Dakiya – Dialog system for email reminders and browsing: In the Dialog Systems lab (Dr. Alex Rudnicky & Dr Alan Black), I am working with my team on developing a dialog system which delivers emails to you (for which you setup reminders) over phone and also allows you to browse your mailbox (Gmail).
- Document classification: In the Language & Statistics course (Dr. Roni Rosenfeld), I worked with my team on developing a classifier for distinguishing real News documents from fake documents (generated by trigram models).

At IIT Kharagpur:-

- Undergraduate Thesis Project – Design and analysis of Gigabit Ethernet switch.
- Distributed hierarchical clustering using MPI java (Parallel Computing course).
- Implementation of a compiler for a C-like language designed for DEC Alpha machines. (Compilers course)
- Augmented the code of NACHOS for implementing virtual, paged memory system (Operating Systems course).

INTERESTING COURSES TAKEN AT CMU

Machine Learning(Dr. Roni Rosenfeld)	Read The Web(Dr. Tom Mitchell)
Information Retrieval(Dr. J. Callan & Dr Y. Yang)	Language and Statistics(Dr. Roni Rosenfeld)
Computational Structural Biology(Dr. C. Langmead)	Algorithms for NLP(Dr. Alon Lavie)
Spoken Dialog Lab(Dr A. Rudnicky & Dr. A. Black)	Grammars and Lexicons(Dr. L. Levin & Dr. T. Mitamura)

COMPUTER SKILLS

Languages: C, C++, JAVA, Perl, Assembly Level Programming, Verilog, TCL/TK
Tools & Packages: Eclipse, WEKA (Machine Learning Package), Minorthird (Information extraction package), Ravenclaw (CMU's framework for complex dialog systems), Tomita V8-4 LR parser (a Natural language parser), CMU NLP toolkit.
Operating Systems: Linux, Unix, MS Windows.

ACADEMIC ACHIEVEMENTS

1. IBM First Patent Application Invention Achievement Award, April 2004.
2. Secured All India Rank 0146 out of 1,20,000 candidates (approx.) in the Indian Institute of Technology's Joint Entrance Examination, one of the toughest undergraduate level admission examination in India.
3. Received National Coal-field Limited meritorious student scholarship for the year 1997-98.
4. Received Birla's best academic performance scholarship in the Sonebhadra region for the year 1995-96.

EXTRACURRICULAR ACTIVITIES

1. Student representative to the faculty for the Master's program 2004-05, currently member of the Student Activities Committee, LTI, Carnegie Mellon University.
2. Public Relations Officer of the Indian Graduate Student Association at CMU.
3. Represent CMU at the Pittsburgh Squash League 'C' and are champions for the past two years: 2004-05, 2005-06. Won the CMU Intra-mural Ultimate Frisbee Championship 2005-06.
4. Coordinator of the "Innovation Curriculum" – an IBM Research worldwide activity during the "Innovation Week-2004" of IBM Research.
5. President of Computer Science & Engineering Dept. Society, I.I.T. Kharagpur, in my Senior Year and held the post of General Secretary in Junior year.
6. Conceptualised *BitWISE*, an international programming contest held by the CSE Department Society, in 2001 and conducted it again in 2002 as the President.
7. Represented IIT Kharagpur in Badminton at Inter IIT sports meet.
8. Captain of R.P. Hall of Residence's badminton team (2002) and a senior member of its Hockey team, currently member of the SCS Intra-mural Floor Hockey team, CMU.
9. Captain of R.P. Hall of Residence's Hindi Dramatics team (2002) and a senior member of English Dramatics & Choreography team.

HOBBIES

- Numismatics
- Playing Squash, Badminton, Hockey, Table Tennis, Volley Ball, Ultimate Frisbee
- Adventure Sports, Hiking
- Dancing

REFERENCES

Available upon request