

## **PALLIKA H KANANI**

Graduate Research Assistant,  
Department of Computer Science, UMass, Amherst

413-695-1399  
[pallika@cs.umass.edu](mailto:pallika@cs.umass.edu)

---

### **EDUCATION:**

- **Ph.D. in Computer Science - Current GPA: 3.6/4.0**  
University of Massachusetts, Amherst, 01/05 – Present  
Advisor: Prof. Andrew McCallum
- **Master of Science in Computer Science - GPA: 3.93/4.0**  
New York University, NY, 09/03
- **Bachelor of Engineering in Computer Science: First Division**  
University of Mumbai, India, 06/00

### **RESEARCH INTERESTS:**

Data mining, Information extraction, Machine learning, Natural language processing.

### **PUBLICATIONS:**

- Kanani, P. and Melville, P., "Prediction-time Active Feature-value Acquisition for Customer Targeting", NIPS 2008 Workshop on Cost Sensitive Learning (Under Review).
- Kanani, P. and McCallum, A., "Efficient Strategies for Improving Partitioning-Based Author Coreference by Incorporating Web Pages as Graph Nodes," Proceedings of the Workshop on Information Integration on the Web (IIWEB 07), pp. 38-43. Also appeared as a poster in NESCAI 2007.
- Culotta, A., Kanani, P., Hall, R., Wick, M. and McCallum, A., "Author Disambiguation using Error-driven Machine Learning with a Ranking Loss Function," a workshop at the Sixth International Workshop on Information Integration on the Web (IIWeb-07, Vancouver, Canada.
- Kanani, P. and McCallum, A., "Resource-bounded Information Gathering for Correlation Clustering," in the Proceedings of Computational Learning Theory 07, Open Problems Track, COLT 2007, LNAI 4539, pp. 625–627, 2007
- Kanani, P., McCallum, A. and Pal, C., "Improving Author Coreference by Resource-bounded Information Gathering from the Web," in the Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 07), pp. 429-434
- Co-authored a poster titled "A microgenetic analysis of the trajectory of motor development" with Dr. Karen Adolph. International Society for Developmental Psychobiology Conference, 2002 and Society for Research in Child Development, 2003.

### **TALKS:**

- "Improving Author Coreference by Resource-bounded Information Gathering from the Web". Indian Institute of Technology, Mumbai, 01/07
- "Resource-bounded Information Gathering from the Web", Machine Learning Friends Lunch, UMass, Amherst, 09/06

### **HONORS and AWARDS:**

- IJCAI'07 student travel award, 01/07
- Verizon Rising Star Fellowship, 09/06.
- Won a second position in a government sponsored knowledge discovery and dissemination challenge for entity resolution task, 08/05

## WORK EXPERIENCE:

- **Graduate Research Assistant, UMass, Amherst (01/05 – Present)**  
I work with Prof. Andrew McCallum in the Information Extraction and Synthesis Lab. My interests include data mining, information extraction, machine learning and NLP. Worked on Author Coreference for scientific paper search engine, Rexa. In many scenarios, functionality or accuracy of machine learning systems can be improved using additional information. My general interest is to study the problem of efficiently querying and incorporating external information under resource constraints.
- **Summer Intern, IBM T.J. Watson Research Center, New York(05/08-08/08)**  
Designed and implemented prediction-time active information acquisition systems for customer targeting. We tested strategies for selecting a subset of instances for which to acquire additional information, so as to get most improvement in accuracy at prediction time with least cost.
- **Project Consultant to White Oak Technologies Inc. Maryland (05/07 – 10/07)**  
Provided consultation for implementing a large-scale system for name disambiguation.
- **Business Development Manager, Neogen Chemicals Ltd. India (10/03 – 12/04)**  
Worked in the family business in various capacities, including IT infrastructure planning and execution, HR policies and administrative activities.
- **Summer Intern, Data Mining Systems Inc. New York (05/03-09/03)**  
Modified a proprietary genetic algorithm application at Deutsche Bank by analysis, design, implementation, testing and evaluation of the required functionality. The new system can analyze time series data where the time horizon is unknown or variable.
- **Software Developer, Psychology Department, NYU (05/02-05/03)**  
Worked on a system to predict “stimulus threshold of infant motor performance” using statistical models. Designed and developed interactive software, including the required database, for special experimental data analysis.
- **Software Developer, Silgate Technologies Pvt. Ltd. India (10/00-09/01)**  
Designed, developed and tested web applications.

## TEACHING EXPERIENCE:

- **Guest Lecture, UMass, Amherst(12/07)**  
Introduction to natural language processing  
Delivered a lecture on Information Extraction.
- **Teaching Assistant, UMass, Amherst (08/06 – 12/06)**  
Introduction to problem solving using computers – Honors section.  
Helped design the course material, graded homework and held office hours.
- **Teaching Assistant, UMass, Amherst (01/06 – 05/06)**  
Computational Linguistics - graded homework and held office hours.
- **Tutor, Foreign Academic Consultancy and Training, India (10/00 – 09/01)**  
Tutored quantitative and analytical skills to 150 students appearing for the GRE, GMAT and SAT tests. Also assisted in creating the course material.

**GRADUATE COURSEWORK:****UMass:**

- Advanced Machine Learning
- Machine Learning
- Advanced Algorithms
- Theory of Computation
- Computer Architecture
- Research Methods

**NYU:**

- Advanced Lab (Advanced topics in Artificial Intelligence)
- Advanced Lab (Research on Data Mining of musical melodies)
- Natural Language Processing
- Machine Learning
- Artificial Intelligence
- Web Search Engines
- Mathematical Statistics
- Advanced Database Systems

**ACADEMIC PROJECTS:**

- Entity Resolution with Spectral methods on Hypergraphs
- Learning to predict Branches. Applied Machine Learning to Branch Prediction.
- Applying spectral clustering to the problem of author coreference.
- Applied data mining techniques on past bond market performance data in order to evaluate the existing financial model for a leading investment analysis firm.
- FindGuru: an information extraction system to mine faculty and students homepages. Used natural language processing to create an interactive database of people in the academia, their research interests and contact information.
- A term paper titled "Knowledge Sharing: from Bibliography to Autonomous Citation Indexing and beyond"
- Implemented a system to extract melodies used for matching against user queries.
- Developed a system for indexing and querying efficiently from web documents.
- Implemented a Naive Bayes language classifier.
- Created a subject focused web crawler.
- Experiments on comparing the relative performance of text mining algorithms.
- Simulated concurrency control and recovery of replicated data.
- Designed a system for planning resource allocation in a manufacturing unit.

**TECHNICAL SKILLS:**

- Languages/Scripting: Java, C/C++, SQL, PL/SQL, PERL, Lisp, Prolog, JavaScript.
- Database/Data Access/Analysis: MATLAB, R, Oracle 8i, MySQL, JDBC, SPSS.

**EXTRA CURRICULAR:**

- Events coordinator for the Indian Students Association, organizing and participating in major cultural events during 2005.
- Organized 'Machine Learning and Friends Lunch' at UMass, Amherst during 2005 -2006.
- Active participant for the annual OISS cultural festival during 2002 and 2003.
- Head of 'Fine Arts Events' in the inter-collegiate cultural festival, Horizons 2000.

**LANGUAGES KNOWN:**

- English, Gujarati, Hindi, Marathi.

**INTERESTS:**

- Music, Classical and Folk Dance, Poetry, Literature, Cooking, Traveling.