

# SHILPA ARORA

EMAIL : SHILPAA@CS.CMU.EDU, HOMEPAGE: [HTTP://WWW.CS.CMU.EDU/~SHILPAA/](http://www.cs.cmu.edu/~shilpaa/)

## OBJECTIVE

---

To pursue research in Natural Language Processing and Machine Learning with a focus on Information Extraction, Question Answering and Active Learning

## EDUCATION

---

- **Ph.D Candidate, Language Technologies Institute**, School of Computer Science, Carnegie Mellon University (CMU), USA ([www.lti.cs.cmu.edu](http://www.lti.cs.cmu.edu)) (CQPA - 3.97/4.33)
- **Masters in Language Technologies** (2008), School of Computer Science, CMU (CQPA - 3.94/4.33)
- **Masters in Computer Science** (2005), Singapore-MIT Alliance - An alliance between Massachusetts Institute of Technology (MIT), National University of Singapore (NUS) and Nanyang Technological University (NTU) (<http://web.mit.edu/sma/>) (CGPA - 4.43/5.0)
- **Bachelors in Computer Engineering** (2004), Nanyang Technological University (NTU), Singapore ([www.ntu.edu.sg](http://www.ntu.edu.sg)). First Class honors (top 5%) (Major GPA - 3.77/4.0)

## AWARDS & ACHIEVEMENTS

---

- **Siebel Scholarship** award for Masters at CMU, 2007-08 (<http://www.siebelscholars.com>)
- **Research Fellowship** for Ph.D & Masters Degree at Carnegie Mellon University, 2006-Present
- **Graduate Fellowship** for Masters Degree with Singapore-MIT Alliance, 2004-05
- **Nanyang Scholarship** for Bachelors Degree at Nanyang Technological University, 2000-04
- **Best Student Poster Award** at Student Summit at DARPA's Machine Reading Conference, 2011
- **Winner in GRID JAVA Developers Challenge** (Buddy-Finder application) organized by Motorola & ST Mobile Data in Singapore, 2003
- Finalist for the **Google Anita Borg Memorial Scholarship**, 2009 ([www.google.com/anitaborg](http://www.google.com/anitaborg))
- Finalist in **Microsoft's .NET development competition**, Imagine Cup Singapore, 2004
- **Dean's Letter of Academic Excellence** (awarded to top 20/440 students), B.E., NTU, 2000-04

## PHD THESIS RESEARCH

---

**Topic:** Minimizing Costs in Generalized Interactive Annotation Learning.

**Proposal:** [http://www.cs.cmu.edu/~shilpaa/ThesisProposal\\_Shilpa\\_Arora.pdf](http://www.cs.cmu.edu/~shilpaa/ThesisProposal_Shilpa_Arora.pdf)

**Main Contributions:**

- **Automatic Feature Engineering:** Generic graph representation for linguistic annotations from which structured features are automatically derived
- **Feature Feedback:** Indirect feedback from the user on structured features
- **Annotation Cost Estimation:** Supervised regression model to estimate annotation cost in a multi annotator environment with feature feedback
- **Joint Selective Sampling:** Interactive learning framework with joint selective sampling of example, annotation strategy and annotator

## SELECTED PUBLICATIONS

---

- Shilpa Arora and Eric Nyberg. *Assessing Benefit from Feature Feedback in Active Learning for Text Classification*. CONLL, 2011.
- Shilpa Arora, Elijah Mayfield, Carolyn P. Rose and Eric Nyberg. *Sentiment Classification using Automatically Extracted Subgraph Features*. Workshop on Emotion in Text at NAACL-HLT, 2010
- Shilpa Arora, Eric Nyberg and Carolyn P. Rose. *Estimating Annotation Cost for Active Learning in a Multi-Annotator Environment*. Active Learning for NLP workshop at NAACL-HLT, 2009
- Shilpa Arora and Eric Nyberg. *Interactive Annotation Learning with Indirect Feature Voting*. Student Research Symposium at NAACL-HLT, 2009
- Shilpa Arora, Mahesh Joshi and Carolyn P. Rose. *Identifying Types of Claims in Online Customer Reviews*. NAACL-HLT, 2009

- Sameer Badaskar, Sachin Agarwal and Shilpa Arora. *Identifying Real or Fake Articles: Towards better Language Modeling*. IJCNLP, 2008
  - Shilpa Arora, Frank Lin, Hideki Shima, Mengqiu Wang, Teruko Mitamura, Eric Nyberg. *Tree Conditional Random Fields for Japanese Semantic Role Labeling*. Unpublished Manuscript, 2008.
  - Sachin Agarwal and Shilpa Arora (2007). *Context Based Word Prediction for Texting Language*. RIAO Conference on Large Scale Semantic Access to Content, 2007.
- (Complete list at <http://www.cs.cmu.edu/~shilpaa/publications.html>)

## WORK EXPERIENCE

### Software Engineering Intern, Google Inc., (May - Aug'11)

*Project: Feature Evaluator for Seti*, a large scale machine learning system developed and used for various learning problems at Google. The evaluator will help to identify potential features and parameter configurations for production by evaluating them on smaller datasets

- Understand the distributed and multi-threaded Infrastructure of Seti, and the existing feature evaluator framework for other machine learning systems
- Integrate Seti into the existing feature evaluator
- Test the evaluator by verifying the past results observed in production
- Work with other teams to use the evaluator for identifying potential configurations for production

### Graduate Research Assistant, Language Technologies Institute, SCS, CMU (Aug'06-Present)

*Projects:*

- **Machine Reading Program (MRP):** "Reading System" that uses information extractors and formal reasoning systems to answer queries from natural language text (Sponsored by DARPA)
  - Integrated and merged the output from several entity and relation extraction components as part of the RACR (Reader And Contextual Reasoner) team on MRP Project - Joint work with IBM Research
- **ROSETTA: Multilingual & Multi-Modal Information Management** (Part of DARPA sponsored GALE program)
  - Developed a UIMA component for preserving source alignment mappings from multiple machine translation engines in MEMT (Multi-Engine Machine Translation)
  - Added functionality to UIMA Component Repository for sharing and validating aggregate components and type systems
- **JAVELIN: Multilingual Question Answering (QA) System** (Sponsored by DTO's AQUAINT program)
  - Developed Tree-CRFs based Semantic Role Labeling (SRL) system for Japanese QA (one of the first work in SRL for Japanese and SRL using Tree-CRFs)
  - Integrated a Chinese SRL system (C-ASSERT) with a faster syntactic parser for Chinese QA

### Analyst , Consulting workforce - Accenture, Singapore (Jun '05-Jun '06)

*Project: DBS Core Banking in India & Indonesia* (Corporate, Enterprise and Consumer banking)

- Designed, deployed and tested the application failover framework.
- Managed the disaster recovery system. Supported the live production system
- Process optimization: Improved and managed the firewall application and tracking process - bottleneck & error prone system. Automated the application deployment procedure
- Worked closely with DBS (client), IBM (client's technology provider) and Infosys (sub-contractor)

### Industrial Attachment, Research Scientist - Singapore Institute of Manufacturing Technology (SIMTech) (Jan- Jun '03)

*Project: Remote operation of an All Terrain Robotic Vehicle (ATRV-2, iRobot)*

- Designed & Developed a Graphical User Interface for the remote operation of a teleoperated vehicle using wireless LAN for video streaming and RF modem for serial port communication

## RELEVANT GRADUATE COURSES & PROJECTS

---

- **Courses:** Machine Learning, Language & Statistics I & II, Algorithms for NLP, Statistical Machine Learning, Software Engineering I & II, Information Extraction, Analysis of Social Media, NLP Lab, Question Answering Lab, Advance Topic in NLP (Opinion & Sentiment Analysis), Artificial Intelligence, Algorithms
- **Projects:**
  - **Open Advancement in Question Answering (OAQA) (Spring '10):** As part of the first OAQA lab (<https://mu.lti.cs.cmu.edu/trac/oaqa/wiki/OAQADocumentation/Introduction>) team at CMU, built an end-to-end framework, wrapped a minimally-configured version of the Ephyra QA system, and improved the system.
    - *Individual contribution:* Improved the existing hybrid **Answer Type classifier** for questions which combines rules with a statistical classifier by choosing the most confident outcome to use rules as soft constraints, i.e., priors on features in the statistical model
  - **Opinion Analysis System (OAS) (Spring '08):** Extended Ephyra QA system to identify opinionated answers from retrieved results

## TEACHING EXPERIENCE

---

**Teaching Assistant, Courses: Software Engineering for Information Systems I & II, Language Technologies Institute, SCS, CMU (Fall '08, Spring '09)**

- Mentored four course projects (SEIS-II): Web recommendation, Machine learning package, Meeting scheduling assistant, Query optimizer
- Created and graded assignments and exams for SEIS-I
- Proposed and formulated a class project on estimating user annotation effort in an Interactive Annotation Learning (IAL) framework for SEIS-I. Developed an end-to-end system with an evaluation framework used by a class of 32 students

## SKILLS AND ABILITIES

---

- **Programming Skills & Applications:** Java, UIMA, C, C++, Perl, latex
- **Machine Learning & Statistics Packages:** Minorthird, SVMLight, Weka, Matlab, R, Mallet Toolkit, etc.
- **Database and Web Skills:** MySQL, Postgres
- **Operating Systems:** Windows, UNIX, Linux, Mac OS
- **Languages:** English (Proficient) and Hindi (Native)

## EXTRA-CURRICULAR ACTIVITIES

---

**Organizer, Student Summit at DARPA's Machine Reading Program Conference, 2011**

- Prepared call for participation and managed reviewing process
- Chaired presentations and group discussion on possible challenge problems for Machine Reading

**Organizer, Opportunities for Undergraduate Research in Computer Science Conference, CMU, 2007** (<http://www.cs.cmu.edu/ourcs/>) Sponsored by Microsoft Research, USA

- Research conference for Undergraduate Women in CS, attended by 60 students internationally
- Organized by CMU for the first time and first of its kind in the US

**Social Director, Organizing Committee for World Universities Debating Championship, Nanyang Technological University (NTU), 2003-04**

- World's largest non-sporting student event - 900 participants from 100 universities in 50 countries
- Received special recognition from the President of Singapore at a tea reception at his residence

**Vice President, Student's Initiative For Technopreneurship (SIFT), NTU, 2002-03**

**REFERENCES:** Available upon request

---