

Nikesh Garera

CONTACT INFORMATION	CSEB 321 Center for Language and Speech Processing Johns Hopkins University Baltimore, MD 21218 USA	<i>Voice:</i> 410.802.5643 <i>E-mail:</i> ngarera@cs.jhu.edu <i>Web:</i> www.cs.jhu.edu/~ngarera
AREAS OF INTEREST	Large-scale information extraction and retrieval, natural language processing, text summarization, machine translation and machine learning. Experience in dealing with large streams of data via MapReduce.	
EDUCATION	Johns Hopkins University , Baltimore, Maryland USA Ph.D. Candidate, Computer Science, Expected graduation date: Sep 2009 (GPA: 3.91/4.0) M.S., Computer Science, Jan 2007 (GPA: 3.94/4.0) Carnegie Mellon University, School of Computer Science , Pittsburgh, Pennsylvania USA M.S., Language Technologies, May 2005 (GPA: 3.90/4.0) University of Mumbai , Mumbai, India B.E., Computer Engineering, 2002 (Graduation class: First class)	
RESEARCH/WORK EXPERIENCE	Johns Hopkins University , Baltimore, Maryland USA Research Assistant (Advisor: D. Yarowsky) Aug., 2005 - Present <i>Center for Language and Speech Processing, Human Language Technology Center of Excellence</i> Research in language technologies including fact extraction, person name disambiguation, coreference resolution and minimally supervised methods for machine translation. Recent Projects: Biographical fact extraction using position-based, transitive and correlational methods. Also developed a method for extracting latent biographical facts using support vector machines exploiting partner-based correlation and non-lexical features. This research also contributed to the cross-document coreference resolution system as part of the JHU-COE team for ACE 2008 competition. Google Inc. , Engineering Intern at Mountain View Campus (Mentor: M. Pasca) Summer 2006 Contributed in developing a large-scale framework for seed-based textual information extraction of class attributes (e.g. <i>make</i> , <i>model</i> , <i>mpg</i> , etc. for the class of <i>cars</i>) from anonymized query logs. Also conducted a preliminary research investigation in building a compare and contrast system by automatically extracting comparative textual segments from the web. Gained extensive experience in dealing with large streams of data via MapReduce. Carnegie Mellon University , Pittsburgh, Pennsylvania USA Research Assistant (Advisor: A. Rudnicky) May 2003 - May 2005 <i>Language Technologies Institute, School of Computer Science</i> Designed and developed a briefing assistant system to facilitate active learning in terms of feature discovery via implicit supervision provided by the end-users using an intuitive highlighting feature of the interface. A briefing/summarization system was developed that learned to identify relevant features of good summary reports by observing content-selection choices that users made while creating the summary during training time. Also, a variant of this system was also deployed for summarizing unstructured interviews with the goal of learning interview-specific summarization behavior over time.	

PUBLICATIONS

N. Garera, D. Yarowsky. Modeling Latent Biographic Attributes in Conversational Genres. To appear in Proceedings of Association for Computational Linguistics (ACL), 2009.

N. Garera, C. Callison-Burch, D. Yarowsky. Improving Translation Lexicon Induction from Monolingual Corpora via Dependency Contexts and Part-of-Speech Equivalences. To appear in Proceedings of the Conference on Natural Language Learning (CoNLL), 2009.

N. Garera and D. Yarowsky. Structural, Transitive and Latent Models for Biographic Fact Extraction. Proceedings of European Chapter of the Association for Computational Linguistics (EACL), 2009.

N. Garera and D. Yarowsky. Minimally Supervised Multilingual Taxonomy and Translation Lexicon Induction. Proceedings of International Joint Conference on Natural Language Processing (IJCNLP), 2008.

N. Garera and D. Yarowsky. Translating Compounds by Learning Component Gloss Translation Models via Multiple Languages. Proceedings of International Joint Conference on Natural Language Processing (IJCNLP), 2008.

M. Pasca, B. V. Durme and N. Garera. The Role of Documents vs. Queries in Extracting Class Attributes from Text. Proceedings of Conference on Information and Knowledge Management (CIKM), 2007.

D. Rao, N. Garera, D. Yarowsky. An Unsupervised Approach to Person Name Disambiguation using Web Snippets. Proceedings of International Workshop on Semantic Evaluations (ACL SemEval), 2007.

M. Kumar, N. Garera, A. I. Rudnicky. Learning from Report-writing Behavior of Individuals. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2007.

N. Garera and D. Yarowsky. Resolving and Generating Definite Anaphora by Modeling Hypernymy using Unlabeled Corpora. Proceedings of the Conference on Natural Language Learning (CoNLL), 2006.

V. Conitzer and N. Garera. Learning Algorithms for Online Principal-Agent Problems (and Selling Goods Online). Proceedings of International Conference on Machine Learning (ICML), 2006.

N. Garera and A. I. Rudnicky. Briefing Assistant: Learning Human Summarization Behavior over Time. AAAI Spring Symposium on Persistent Assistants, 2005.

N. Garera (Advisors: S. Vogel and R. E. Frederking). Estimating Better Position Alignments for Machine Translation. Tech Report at Carnegie Mellon University.

N. Garera, D. Modi, A. Shah, N. Valia (Advisor: A. S. Moharir). Web security using Role Based Access Control (RBAC). Undergraduate thesis project conducted at IIT Bombay.

J. Mayfield, D. Alexander, B. Dorr, J. Eisner, T. Elsayed, T. Finin, C. Fink, M. Freedman, N. Garera, P. McNamee, S. Mohammad, D. Oard, C. Piatko, A. Sayeed, Z. Syed, R. Weischedel, Cross-Document Coreference Resolution: A Key Technology for Learning by Reading. AAAI 2009 Spring Symposium on Learning by Reading and Learning to Read, March 2009.

PROFESSIONAL ACTIVITIES AND SERVICE

Program committee member for Empirical Methods in Natural Language Processing (EMNLP) 2009
Program Committee member for European Chapter of Association for Computational Linguistics (EACL) 2009

Teaching Assistant for Information Extraction Laboratory at JHU Summer School on Human Language Technology, Jul 2007.

Student Volunteer for International Conference on Machine Learning (ICML), 2006

Member of Association for Computational Linguistics (ACL)

Member of Association for the Advancement of Artificial Intelligence (AAAI)

INVITED TALKS Translating Compounds by Learning Component Gloss Translation Models via Multiple Languages. IIT Bombay, Mumbai, India, Jan 2008.

 Towards a Personal Briefing Assistant, LTI Student Research Symposium, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, Sep 2004.

HONORS/AWARDS Recipient of Ambassador of T.S.E.C award for all round excellent performance throughout four years of undergraduate program.

 Recipient of junior college and government merit award (from Municipal Corporation of Navi Mumbai) for standing second in my college at the H.S.C board examination.

COMPUTER SKILLS Languages: C, C++, Perl (also some exposure to Java, Python, SQL)
 Software Packages: Matlab, R, SVM-Light, Weka, Minorthird, SRILM

RELEVANT
GRADUATE LEVEL Natural Language Processing, Information Retrieval, Information Extraction, Machine Learning,
COURSES Intermediate Statistics, Language and Statistics, Data Structures and Algorithms, Randomized Algorithms, Information Theory, Foundations of Optimization, Machine Learning Theory.

REFERENCES Available on Request.