CURRICULUM vitae

Nitin MADNANI

Department of Computer Science
University of Maryland
6922 Hanover Pkwy
Greenbelt MD 20770
Cell: (301) 706 - 3539

E-mail: nmadnani@umiacs.umd.edu
http://www.umiacs.umd.edu/~nmadnani
Current Status: Student (F-1) Visa
Citizenship: India

Education

University of Maryland, College Park, MD.
Ph.D. in Computer Science, GPA: 4.0
Dissertation Topic: Automatic Paraphrase Generation for Natural Language Processing
expected May 2010.

University of Maryland, College Park, MD.
M.S in Computer Engineering, GPA: 3.77
Concentration: Computer Organization, Microrchitecture and Embedded Systems
2004.

Punjab Engineering College, Panjab University, India.
B.E. in Electrical Engineering, With Honors
Senior Thesis: Interactive Visualization of Grounding Systems for Power Stations
2000.

Professional Interests

Computational Linguistics, Natural Language Processing, Statistical Machine Translation, Automatic Paraphrase Generation, Machine Learning, Artificial Intelligence and Computer Science Education.

Teaching Experience

Instructor, Computational Linguistics I, Department of Computer Science, University of Maryland (Fall 2007, Fall 2008)

- Redesigned the entire course curriculum to cater to a diverse audience (both **graduate** and **undergraduate** students from Computer Science & Linguistics). It has been used as the default curriculum for every iteration of the course since then.
- Designed and delivered lectures on several language processing topics such as Part-of-Speech Tagging, Hidden Markov Models, Expectation Maximization, N-gram Language Modeling, Non-CFG Parsing Models.
- Developed and created homework assignments and programming projects using Python and NLTK (Natural Language Toolkit) that allowed students to imbibe course concepts in a hands-on fashion.
- Created and monitored an online forum to answer students' questions promptly and in detail.
- Guided the assigned teaching assistant(s) on how to grade homeworks and projects.
- Received extremely positive ratings from the students at the end of the semester. Students really appreciated the curriculum design, the hands-on instructive quality of assignments and the one-on-one attention via the forum and the office hours.

Guest Lecturer, Computational Linguistics I, Department of Computer Science, University of Maryland, Fall 2009

- Designed and conducted a hands-on session to introduce students to real-world language processing using presidential state of the union addresses and congressional floor debates.

Teaching Assistant, Introduction to Natural Language Processing, Department of Computer Science, University of Maryland, Fall 2004

- Graded homeworks and projects.
- Held regular office hours to help students with homeworks, projects and the course material, in general.

Private Tutor, Department of Computer Science, University of Maryland, 2002–2003

- Tutored undergraduate students having difficulty with the Operating Systems course offered by the department.
- Helped students understand the core concepts in the lectures and apply them to the programming projects.

Computer Science Instructor, Yantra Shikshan Foundation, Delhi, India, 2000–2001

- Volunteered with the Yantra Shikshan Foundation, an Indian NGO dedicated to teaching basic Computer Science to students from rural and impoverished areas.
- Used Squeak/Etoys and Linux to encourage elementary and high school students to explore and understand CS fundamentals on their own.
- Part of committee tasked with developing an atypical CS curriculum aimed at such students.
- Taught basic Linux System Administration to students from local vocational institutions.

Teaching Assistant, Digital Computations, Department of Electrical Engineering, Punjab Engineering College, 1998

- Conducted regular tutorials to teach students how to program in FORTRAN and C.
- Graded assignments and projects.
- Supervised laboratory sessions.

Upcoming Teaching

Instructor, Introduction to Cloud Computing, Department of Computer Science & iSchool, University of Maryland (Spring 2010)

Research Experience

Research Assistant, University of Maryland Institute for Advanced Computer Studies, Laboratory for Computational Linguistics & Information Processing, 2004–Present

Sentential Paraphrase Generation

- Designed, developed and implemented a novel, feature-driven computational model for automatically paraphrasing any given sentence in any language to another semantically equivalent sentence. The model is particularly appropriate for use in other language processing applications (see **Machine Translation** below).
- Conducting human studies using Amazon Mechanical Turk in order to understand how humans perceive semantic equivalence at the sentence level and to explore how this perception compares with the computational model.

Machine Translation

- Participated in the development of a rule-based framework (`DUSTer`) to unravel cross-linguistic *divergences*—naturally occurring instances wherein the same underlying concept is distributed over different words between two natural languages—that can confound the process of automatic translation.
- Ported `DUSTer` to an entirely new language pair (Hindi-English) as part of the DARPA Translingual Information Detection, Extraction and Summarization (TIDES) program.

- Ported a popular machine learning algorithm used to automatically learn the system parameters of a statistical machine translation system from Perl to C using the `Inline::C` perl module. The implementation was used in a state-of-the-art translation system that was ranked highly at the Annual NIST Machine Translation Evaluation in 2005.
- Integrated sentential paraphrasing model (described above) with a state-of-the-art machine translation system to solve a significant research problem in current translation methods: the requirement of *multiple* reference translations for automatically learning the system parameters using the algorithm above. Use of said model led to a statistically significant, empirically verified gain in system performance.
- Designed and co-authored a state-of-the-art metric (`TERp`), written in `Java`, used by the research community to evaluate output of machine translation systems. `TERp` improves upon a previously existing metric by incorporating semantic enhancements like synonyms and paraphrases. The metric was judged to be the top-performing metric at the NIST Machine Translation Metrics Challenge in 2008.
- Developing next generation machine translation systems for the DARPA Global Autonomous Language Exploitation (GALE) program as a member of a research consortium led by BBN Technologies.

Automated Text Summarization

- Refined an existing multi-document summarization system (`TRIMMER`) that generates summaries by extracting relevant sentences and compressing them. The main refinement was implementing a novel method for automatically finding the set of feature weights that maximize system performance and was ranked 2^{nd} at the Document Understanding Conference (DUC) organized by NIST in 2007.
- Redesigned and reimplemented `TRIMMER` so that it could be run more efficiently on a 20-node PBS cluster.
- Determining the right order for sentences in a multi-document summary is a non-trivial problem. Conducted extensive human studies to better understand how sentences in a summary should be ordered so as to maximize its coherence.

Information Retrieval

- Developed and implemented a simulated interactive question answering system to understand how introducing elements of interaction, such as clarification questions, can improve retrieval performance. The system was evaluated in the ciQA (complex interactive question answering) task at the Text Retrieval Conference organized by NIST in 2007.

Text Visualization

- Developed and implemented the first prototype of `EMILY`, a tool for visualization and analysis of Emily Dickinson's poetry. The tool was developed in collaboration with and the Human Computer Interaction Lab (HCIL) and Maryland Institute for Technology in the Humanities (MITH). It has since been developed further and is being used by humanists in several institutions.

Research Intern, IBM T J Watson Research Center, Natural Language Group, Summer 2005

- Worked on the **MALACH** (Multilingual Access to Large Spoken Archives) project aimed at automatically extracting information from 116,000 hours of digitized interviews in 32 languages from 52,000 survivors, liberators, rescuers and witnesses of the Nazi Holocaust. The extraction system is trained on parts of archives that have been manually annotated.

- Developed and tested active learning strategies to improve the performance of the information extraction system. The best strategy reduces the human annotation required by 50% without affecting the system performance.

Research Intern, Embedded Systems Group, Netrino LLC, Summer 2002

- Wrote Application and System software for the TERN A-CORE board with the AMD188ES microprocessor, and for the low power ECOG1 Microcontroller—which was used in "Embedded Programming 101" at the Embedded Systems' Conference held in 2002.
- Ported Quantum Framework—a C++ framework to program embedded systems using UML Statecharts—to `Micro/C-OS2`, a real-time preemptive kernel.
- Designed and coded a temperature-estimating cricket emulator that varies its chirp rate according to the microprocessor core temperature, on the Cyan Technologies' low power communication processor (ECOG1). Wrote device drivers for the analogue speaker interface.
- Conducted background research for *Embedded Systems Dictionary*, published by CMP Books in 2003 (ISBN: 15782012090).

Publications

Journals

- Matthew Snover, **Nitin Madnani**, Bonnie Dorr and Richard Schwartz. TER-Plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate. *Machine Translation Journal, Special Issue on: Automated Metrics for MT Evaluation*, In Press.
- **Nitin Madnani**. Querying and Serving N-gram Language Models with Python. *The Python Papers*, 4(2). 2009.
- **Nitin Madnani**. Source Code: Querying and Serving N-gram Language Models with Python. *The Python Papers Source Codes*, 1(1), 2009.
- **Nitin Madnani**. Getting Started on Natural Language Processing with Python. *ACM Crossroads*, 13(4), 2007.
- Bonnie J. Dorr, Necip Fazil Ayan, Nizar Habash, **Nitin Madnani**, and Rebecca Hwa. Rapid Porting of DUSTer to Hindi. *ACM Transactions on Asian Language Information Processing*, 2(2), 2003.

Conferences

- **Nitin Madnani** and Jimmy Lin. The Python and The Elephant: Large Scale Natural Language Processing with NLTK and Dumbo. *Proceedings of the Eighth Annual Python Conference (PyCon)*, To appear in 2010.
- **Nitin Madnani**, Philip Resnik, Bonnie Dorr and Richard Schwartz. Are Multiple Reference Translations Necessary? Investigating the Value of Paraphrased Reference Translations in Parameter Optimization. *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas (AMTA)*, 2008.
- Saif Mohammad, Bonnie J. Dorr, Melissa Egan, **Nitin Madnani**, David Zajic, and Jimmy Lin. Multiple Alternative Sentence Compressions and Word-Pair Antonymy for Automatic Text Summarization and Recognizing Textual Entailment. *Proceedings of the Text Analysis Conference (TAC)*, 2008.
- **Nitin Madnani**, Jimmy Lin, and Bonnie Dorr. TREC 2007 ciQA Task: University of Maryland. *Proceedings of the Sixteenth Text Retrieval Conference (TREC)*, 2007.
- **Nitin Madnani**, David Zajic, Bonnie Dorr, Necip Fazil Ayan and Jimmy Lin. Multiple Alternative Sentence Compressions for Automatic Text Summarization. *Proceedings of the Document Understanding Conference (DUC)*, 2007.

- David Chiang, Adam Lopez, **Nitin Madnani**, Christof Monz, Philip Resnik and Michael Subotin. The Hiero Machine Translation System: Extensions, Evaluation, and Analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 2005.

Workshops

- Matthew Snover, **Nitin Madnani**, Bonnie Dorr and Richard Schwartz. Fluency, Adequacy, or HTER? Exploring Different Human Judgments with a Tunable MT Metric. *Proceedings of the Fourth ACL Workshop on Statistical Machine Translation (WMT)*, 2009.
- Matthew Snover, **Nitin Madnani**, Bonnie Dorr and Richard Schwartz. TERp: A System Description. *Proceedings of the First NIST Metrics for Machine Translation Challenge (MetricsMATR)*, 2009.
- **Nitin Madnani** and Bonnie Dorr. Combining Open-Source with Research to Re-engineer a Hands-on Introductory NLP Course. *Proceedings of the Third ACL Workshop on Issues in Teaching Computational Linguistics (TeachCL)*, 2008.
- **Nitin Madnani**, Necip Fazil Ayan, Philip Resnik, Bonnie Dorr. Using Paraphrases for Parameter Tuning in Statistical Machine Translation. *Proceedings of the Second ACL Workshop on Statistical Machine Translation (WMT)*, 2007.
- **Nitin Madnani**, Rebecca Passonneau, John Conroy, Necip Fazil Ayan, Bonnie Dorr, Judith Klavans, Dianne O'Leary and Judith Schlesinger. Measuring Variability in Sentence Ordering for News Summarization. *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG)*, 2007.

Posters

- **Nitin Madnani**, Philip Resnik, Bonnie Dorr and Richard Schwartz. Applying Automatically Generated Semantic Knowledge: A Case Study in Machine Translation. *Proceedings of the NSF Symposium on Semantic Knowledge Discovery, Organization and Use*, 2008.
- Catherine Plaisant, **Nitin Madnani**, Matt Kirschenbaum, Martha Nell Smith, Tanya Clement and Greg Lord. Exploring Emily Dickinson Letters. *Proceedings of the 22nd Annual Human-Computer Interaction Lab Symposium*, University of Maryland, 2005.
- **Nitin Madnani**, Necip Fazil Ayan, Bonnie Dorr, Nizar Habash, Christof Monz. Portable Divergence Unraveling: The Case of Hindi. *Research Review Day*, University of Maryland, 2004.

Working Papers

- Generating Phrasal & Sentential Paraphrases: A Survey of Data-Driven Methods.
  Journal article in second round of review for Computational Linguistics (MIT Press).
- Machine Translation Evaluation and Optimization.
  Book chapter in preparation.
- A Pythonic Exploration of Vector Space Methods for Semantic Similarity.
  Article in preparation.

Unpublished Manuscripts

- EMILY: A Tool for Visual Poetry Analysis, 2005.
- Active Learning for Mention Detection: A Comparison of Sentence Selection Strategies, 2005.

Selected Oral Presentations

A Summer of (Active) Learning, Computational Linguistics Colloquium. University of Maryland, College Park, MD, September 2005.

Multiple Alternative Sentence Compressions for Automatic Text Summarization. Document Understanding Conference. Rochester, NY, April 2007.

Measuring Variability in Sentence Ordering for News Summarization. European Workshop on Natural Language Generation. Schloss Dagstuhl, Germany, June 2007.

Using Paraphrases for Parameter Tuning in Statistical Machine Translation. Workshop on Statistical Machine Translation. Annual Meeting of the Association for Computational Linguistics. Prague, Czech Republic, June 2007.

Using Paraphrases for Parameter Tuning in Statistical Machine Translation. Invited talk. Annual Technical Meeting for Global Autonomous Language Exploitation. San Francisco, CA, June 2007.

Using Open-Source and Research to Re-engineer a Hands-on Introductory NLP Course. Workshop on Issues in Teaching Computational Linguistics. Annual Meeting of the Association for Computational Linguistics. Columbus, OH, June 2008.

Investigating the Value of Paraphrased Reference Translations in Parameter Optimization. Conference of the Association for Machine Translation in the Americas. Waikiki, Hawaii, October 2008.

Applying Automatically Generated Semantic Knowledge: A Case Study in Machine Translation. NSF Symposium on Semantic Knowledge Discovery, Organization and Use. New York University, NY, November 2008.

Open Source Software

Developer & Project Member, *Natural Language Toolkit*, `http://www.nltk.org`.

A community driven suite of Python modules, data and documentation for research and development in natural language processing. Personal contributions include development of new modules, inclusion of new data and several bug fixes and improvements. NLTK is widely used in pedagogy and a list of courses using it can be found at `http://www.nltk.org/courses`

Primary Developer. *Scripting Language Model Interface*.

A general purpose interface to a popular language modeling toolkit (SRILM) that allows reading and querying these language models *directly* in Python, Perl and most other scripting languages. In use by NLP research groups at University of Illinois, Simon Fraser University and Institute for Mathematical Sciences (India).

Primary Developer. *Light-weight Language Model Server*.

A Python-based XML-RPC server for language models that allows multiple clients to query the *same* language model loaded in server mode.

Developer, *UMIACS Word Alignment Interface*.

A Java-based tool for creating and viewing word alignments between language pairs. It has been widely used across the community to create alignments for many language pairs including Hindi-English, Welsh-English, Swahili-English, Czech-English and Chinese-English.

Principal Service Activities

Peer Mentor, Department of Computer Science, University of Maryland, 2007–2008

Referee

- *Journal of Machine Translation*.
- Squibs and Discussions, *Computational Linguistics*.

Program Committee Member

- *International Conference on Computational Linguistics (COLING)*, 2010.
- *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2009.
- *Conference of the European Chapter for the Association for Computational Linguistics (EACL)*, 2009.
- *Conference of the North American Chapter for the Association for Computational Linguistics (NAACL)*, 2009

Reviewer

- *Conference of the Association for Machine Translation in the Americas (AMTA)*, 2008
- *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2007
- *International Joint Conference on Natural Language Processing (IJCNLP)*, 2005

Awards & Scholarships

National Talent Search Examination Award, India, 1993

Merit Scholarship, Punjab Engineering College, Panjab University, 1997–2000

Graduate Research Assistantship, Institute for Advanced Computer Studies, University of Maryland, 2004–Present

Jacob K. Goldhaber Travel Grant, University of Maryland, 2005

Professional Affiliations

Association for Computing Machinery (ACM)

ACM Computer Science Teachers Association (CSTA)

Association for Computational Linguistics (ACL)

ACL Special Interest Group on Machine Translation (SIGMT)

ACL Special Interest Group on Natural Language Generation (SIGGEN)

ACL Special Interest Group on Computational Semantics (SIGSEM)

References

Bonnie Dorr
Professor
Department of Computer Science
University of Maryland
College Park, MD 20742

Philip Resnik
Associate Professor
Department of Linguistics
University of Maryland
College Park, MD 20742

Jimmy Lin
Associate Professor
College of Information Studies
University of Maryland
College Park, MD 20742

Chris Callison-Burch
Assistant Professor
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218

Skills

C/C++, Java, LaTeX, Matlab, Perl, Python, R, Ruby.

Unix, Linux, MS-DOS, MS-Windows, Mac OS X.

Fluent spoken/written English, Hindi; fair spoken Punjabi and Sindhi.