

Curriculum Vitae

Mahesh Joshi

E-Mail: maheshj@cmu.edu

Web: <http://www.d.umn.edu/~joshi031/>

Education

- August 2006 – present: Masters in Language Technologies, Carnegie Mellon University
- September 2004 – August 2006: MS, University of Minnesota Duluth
 - Major: Computer Science
 - Minor: Statistics
 - GPA: 4.0
- July 1997 – June 2001: Bachelor of Computer Engineering, Government College of Engineering (now Pune Institute of Engineering and Technology), Pune, India
 - Major: Computer Engineering
 - Aggregate percentage: 66.83 (First Class with Distinction)

Research Experience

- **Masters' thesis, University of Minnesota, Duluth (August 2006):** My research focused upon developing special purpose kernels for support vector machines applied to the task of word sense disambiguation and acronym expansion in medical text. I have applied semantic kernels (similarity matrices) learned from unlabeled text to the problems of word sense disambiguation and abbreviation expansion. Please see <http://www.d.umn.edu/~joshi031/research/research.html> for more details.
- **Research Assistant to Dr. Ted Pedersen at University of Minnesota, Duluth (May 2006 – July 2006):** I worked on adding support for Latent Semantic Analysis (LSA) based data representation in the unsupervised clustering package SenseClusters (<http://senseclusters.sourceforge.net/>) for performing feature clustering and LSA based context clustering.
- **Summer 2005 Internship at the Mayo Clinic, Division of Biomedical Informatics:** My research involved automatically deciding the correct expansion of ambiguous acronyms in medical text. I worked on identifying features useful for this problem and applying supervised machine learning methods to this problem. A short report of my internship can be viewed at <http://www.d.umn.edu/~joshi031/files/summer2005.txt> and the final presentation slides are available at <http://www.d.umn.edu/~joshi031/files/summer2005.pdf>.

Publications

- Mahesh Joshi, Serguei Pakhomov, Ted Pedersen, Christopher Chute. "A Comparative Study of Supervised Learning as Applied to Acronym Expansion in Clinical Reports." *To appear in Proceedings of the American Medical Informatics Association Annual Symposium (AMIA-2006)*.
<http://www.d.umn.edu/~joshi031/files/AMIA06JoshiM.pdf>
- Mahesh Joshi. "Kernel Methods for Word Sense Disambiguation and Abbreviation

Expansion in the Medical Domain.” *Masters' Thesis, University of Minnesota, Duluth. August 2006.*

<http://www.d.umn.edu/~joshi031/files/MSThesis-JoshiM.pdf>

- Mahesh Joshi, Serguei Pakhomov, Ted Pedersen, Richard Maclin and Christopher Chute. “An End-to-end Supervised Target-Word Sense Disambiguation System.” *Appears in Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-2006), Intelligent Systems Demonstrations.*
<http://www.d.umn.edu/~joshi031/files/AAAI0603JoshiM.pdf>
- Mahesh Joshi, Ted Pedersen, Richard Maclin, Serguei Pakhomov. “Kernel Methods for Word Sense Disambiguation and Acronym Expansion.” *Appears in Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-2006), Student Abstract Papers.*
<http://www.d.umn.edu/~joshi031/files/AAAI0612JoshiM.pdf>
- Mahesh Joshi, Ted Pedersen and Richard Maclin. “A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain.” *Appears in Proceedings of the Second Indian International Conference on Artificial Intelligence (IICAI 2005), Pune, India.* Available at
<http://www.d.umn.edu/~joshi031/files/iicai2005-wsd.pdf>

Research and Class Presentations

- A Comparative Study of Support Vector Machines Applied to the Supervised Word Sense Disambiguation Problem in the Medical Domain – Dec. 20th 2005. Talk presented at the Second Indian International Conference on Artificial Intelligence (IICAI 2005), Pune, India on Dec 20th 2005.
- Discussion in the NLP@UMD group about Domain Kernels for Word Sense Disambiguation (<http://acl.ldc.upenn.edu/P/P05/P05-1050.pdf>), a paper from ACL 2005 by Alfio Gliozzo, Claudio Guiliano and Carlo Strapparava – Oct. 14th 2005.
Slides: <http://www.d.umn.edu/~joshi031/files/domain-kernels.pdf>
- Seminar at the Mayo Clinic, Division of Biomedical Informatics - presentation about summer internship work – Aug. 25th 2005.
Slides: <http://www.d.umn.edu/~joshi031/files/summer2005.pdf>
- Graduate class presentation on Grid Computing, for Advanced Computer Architecture class – Apr. 28th 2005.
Slides: <http://www.d.umn.edu/~joshi031/files/grid-computing.pdf>
Relevant extracts from referenced documents:
<http://www.d.umn.edu/~joshi031/files/grid-computing.txt>

Research Software Released

- **NSPGate** (<http://nspgate.sourceforge.net/>): A GATE (<http://gate.ac.uk/>) plug-in for the Ngram Statistics Package (NSP - <http://ngram.sourceforge.net/>). Provides a wrapper for NSP, to mark-up n-gram annotations in GATE documents. Version 0.03 (<http://sourceforge.net/projects/nspgate/>) released on July 15th 2006.
- **WSDGate** (<http://wsdgate.sourceforge.net/>): A word sense disambiguation package based on GATE, NSP, NSPGate and WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>). Version 0.05 (<http://sourceforge.net/projects/wsdgate/>) released on July 15th 2006.
- **WSDShell**: A word sense disambiguation toolkit that makes use of NSP, SenseTools and

WEKA to run WSD experiments on Senseval-2 formatted data. Version 0.05 (<http://www.d.umn.edu/~joshi031/files/wsdshell/wsdshell-v0.05.tar.gz>) released on Dec. 11th 2005.

- **nlm2sval2**: A tool to convert the National Library of Medicine WSD collection (<http://wsd.nlm.nih.gov/>) into Senseval-2 (<http://www.d.umn.edu/~tpederse/Code/Readme.SenseTools-0.5.intro.txt>) format. Version 0.02 (<http://www.d.umn.edu/~joshi031/files/nlm2sval2/nlm2sval2-v0.02.tar.gz>) released on Mar. 3rd 2005.

Academic Appointments

- **Graduate Research Assistant**
Visualization and Digital Imaging Lab
University of Minnesota Duluth
August 2005 – present
 - Responsibilities include:
 - Assisting the faculty researchers using the laboratory
 - Software development of projects done at the laboratory
 - Maintenance of laboratory hardware and software resources
 - Website maintenance and update
- **Graduate Teaching Assistant**
University of Minnesota Duluth
September 2004 – May 2005
 - With Dr. Richard Maclin (CS 4611 - Database Management Systems), Dr. Ted Pedersen (CS 3221 - Operating Systems Practicum) and Dr. James Allert (CS 1121 - Visual Basic .NET). Please refer to <http://www.d.umn.edu/~joshi031/ta/teaching.html> for more details.

Academic Projects

- **Empirical Comparison of Machine Learning Algorithms using a Bias-Variance Measure (Fall 2005)**: This graduate class project involves adapting the implementation of Bias-Variance decomposition (Ron Kohavi and David Wolpert – “Bias Plus Variance Decomposition for Zero-One Loss Functions”, *Machine Learning, Proceedings of the Thirteenth International Conference, 1996*) available in WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) for the purpose of empirical comparison of machine learning algorithms implemented in WEKA, using the UCI datasets (<http://www.ics.uci.edu/~mlearn/MLRepository.html>).
- **Simple Essay Analysis Mechanism - <http://seam.sourceforge.net/> (Fall 2004)**: This group project was implemented as a part of the Natural Language Processing graduate class. The project aims at automated evaluation of student essay responses in competitive examinations like GRE and TOEFL. It focuses on the aspects of relevance, gibberish detection, fact identification and fact verification. I was involved in the relevance analysis module that deals with assigning a relevance score to the essay based on its similarity to the essay prompt. We used Latent Semantic Analysis (LSA) methods for similarity measurement.
- **Network Intrusion Detection System (July 2000 - June 2001)**: This group project was implemented during my senior year of Bachelor of Computer Engineering. We developed a

Linux-based system for intrusion detection. It consists of two main sub-systems – the “network-based” sub-system and the “host-based” sub-system. The network-based sub-system involves scanning of network packets and detection of malicious attacks using known attack signatures. The host-based sub-system consists of modification to the Linux kernel to add additional security to critical APIs. This project won the 1st prize at the CSA Showcase 2001 competition held at the Indian Institute of Science, Bangalore. See <http://www.d.umn.edu/~joshi031/acad/nids.html> for the README and source code.

Work Experience

- **Member of Technical Staff**, Persistent Systems Pvt. Ltd., Pune, India. July 2001 – July 2004
 - Worked as a part of Data Management team for Agilent Cerity Network Data System (<http://www.agilent.com> > Products & Services > Life Sciences/Chemical Analysis > Data Systems > Cerity Pharmaceutical for QA/QC)
 - Responsibilities included design and development of Data Archive and Restore utility, Database Qualification utility, development of Query module and maintenance of code related to Database Abstraction Layer. Was leading a sub-team of 2 engineers for the last six months of job duration.

Technical Skills

- **Machine Learning and NLP Tools:** WEKA, Ngram Statistics Package, GATE
- **Programming languages:** C, C++, Java, Perl, Visual Basic 6.0, Visual Basic .NET
- **Technologies:** COM/DCOM
- **Platforms:** Linux, Macintosh, Solaris, Windows
- **Tools:** Merant PVCS Tracker, Numega BoundsChecker, Rational ClearQuest
- **Version control systems:** CVS, Rational ClearCase
- **Databases:** Oracle (some experience in PL/SQL)

Scholarships and Assistantships Awarded

- Graduate Research Assistantship and Full Tuition Waiver (2005 – 2006)
- Summer Award from Dr. Ted Pedersen, Associate Professor, Department of Computer Science (2005)
- Graduate Teaching Assistantship and Full Tuition Waiver (2004 – 2005)
- National Talent Search Scholarship (1995 – 2001)
- Dhirubhai Ambani Foundation Undergraduate Scholarship (1997 – 2001)
- National Merit Scholarship (1997-98)
- Maharashtra State Merit Scholarship (1997-98)

Academic Honors

- 1st prize at CSA Showcase 2001, Indian Institute of Science, Bangalore, India for the senior year project “Network Intrusion Detection System” (2001)
- Ranked 2nd among more than 100,000 students in Higher Secondary Examination (1997)
- 1st rank in Physics (100/100) at Higher Secondary Examination (1997)

- Ranked 23rd among more than 125,000 students in Secondary School Examination (1995)

Certifications and Memberships

- Grade A qualification of Business English Certificate, Level 2, University of Cambridge, London
- Life member of [MENSA](http://in.mensa.org/index.html) (<http://in.mensa.org/index.html>), India

References

- Dr. Richard Maclin
Associate Professor
Department of Computer Science, University of Minnesota Duluth
rmaclin@d.umn.edu
- Dr. Ted Pedersen
Associate Professor
Department of Computer Science, University of Minnesota Duluth
tpederse@d.umn.edu
- Dr. Serguei Pakhomov
Assistant Professor of Biomedical Informatics
Mayo Clinic College of Medicine, Rochester
Pakhomov.Serguei@mayo.edu