University of Massachusetts Amherst Fall/Winter 2008

Significant BITS

Newsletter of the
Department of Computer Science

RIPPLES – a MANIC decade

Professor Rick Adrion

The Research in Presentation Production for Learning Electronically Group (RIPPLES) has seen more than a decade of truly *MANIC* research, development, and educational applications. This research has been driven by rapidly changing technologies, creative new teaching and learning approaches, student and curricula demands, and interdisciplinary research. Following the development of the Multimedia Asynchronous Networked Individualized Courseware (MANIC) system by Distinguished Professor Jim Kurose and several graduate and undergraduate students, the RIPPLES Group and its predecessor, the Advanced Multimedia Laboratory (AML), have focused on how to most effectively use World Wide Web and CD/DVD technologies to deliver lectures and course materials in and outside of the classroom.

RIPPLES, directed by Computer Science Professors Rick Adrion, Kurose and Engineering Professor Wayne Burleson has gone through five distinct phases: streaming multimedia courseware as a "driving application" for network research; a period of experimentation with intelligent tutors, authoring tools, embedded assessment and integration with classroom pedagogy; CD/DVD-based media coupled with internet support in a production environment; extensions to support constructivist pedagogies in an open source delivery system; and automated content capture.

The AML/RIPPLES group developed a variety of "record-and-playback" multimedia content delivery systems using the MANIC framework. Content is delivered in a multimedia format that includes synchronized video (typically of an instructor), audio, and classroom materials (text, graphics, screen capture, whiteboard images, etc.) and an index (text and/or image thumbnails). The index provides direct access and playback to any portion of the content and supplements the linear play/forward/rewind controls. Various MANIC content delivery systems include functions to support navigation, search, collaboration, annotation, and user-controlled views.

Beginning in 1995, with funding from the National Science Foundation, Kurose led a research effort to investigate networked multimedia information systems that involved CS professors Jack Stankovic (now at U. Virginia), Don Towsley, Krithi Ramamritham (now at IIT Mumbai), Eliot Moss, Bruce Croft, Kathryn McKinley (now at Texas Austin), Adrion, and several others. To evaluate multimedia protocols for a continuous-media courseware server, Kurose and his students developed the MANIC system, which was to become the framework for research on learning systems, a test bed for studies on the interdependence of technology and pedagogy in on- and off-campus education, and the delivery system for a significant number of distance education courses.

continued on page 4

Banner year for faculty

From Lifetime Achievement Awards to new faculty NSF CAREER awards, our faculty have received many honors this year. Read more about the specifics in this issue.

In Memoriam
Professor Paul Utgoff 8

News
Kurose named NSM Dean 7

Awards
Diao & Wang receive CAREER awards 5
SIGCOMM honors Towsley 6

Alumni
Recent Ph.D. graduates 16

Department welcomes four new faculty members

The Department of Computer Science has attracted four strong new hires through the faculty recruiting efforts. Andrew McGregor and Yannis Smaragdakis are joining the faculty as tenure-track faculty, and David Smith and Michael Zink are new research-track faculty. The new faculty have research strengths in applied programming languages and software engineering, algorithms and complexity, machine learning and natural language processing, and distributed systems/computer networks.

continued on page 2

From the Chair

Andrew Barto

Letters from the chair are not usually about technical subjects, but as I write this I am having trouble ignoring the uncertainty we are facing both locally and around the globe. With so many researchers in Computer Science and other disciplines studying decision making under uncertainty, one might think that we would be able to turn the crank on any number of sophisticated algorithms to figure out what is the best course of action for maximizing our 'expected utility' even while lacking complete information about the state of the world, about how our decisions might change things, and about what the unfolding future will mean for our measures of utility.

Why aren't decision theorists saving the day? Indeed, why has some of the blame for the current financial crisis been attributed to new financial instruments that owe their existence to complex theories and algorithms? There are many answers to these questions, but putting aside the big issue of how we assign utility to possible futures, the main answer is that these methods need either accurate probabilistic models or immense amounts of data, not to mention a lot of regularity in what is happening in the world. Lacking these, we can argue about whose probability estimates are the best, with our inference algorithms waiting in the wings to grind out recommendations, all the while knowing that there is no infallible way to predict the future.

But evolution has given us skills for dealing with uncertainty—not skills that necessarily help us manage our finances, but skills that have kept our ancestors alive through the millennia. It is not known exactly how we are wired to act in the face of uncertainty, but a few theories—even if false—are thought provoking. One is that the brain has different systems, mediated by different neurotransmitters, for dealing with expected uncertainty and unexpected uncertainty (A. J. Yu and P. Dayan, "Uncertainty, Neuromodulation, and Attention," *Neuron*, 46, 681-692, 2005). Expected uncertainty occurs when we are in a familiar environment and have had a lot of experience with its unreliable predictive relationships. We have pretty good probability estimates and strong top-down expectations. Unexpected uncertainty, on the other hand, occurs when there is a major shift in context causing our top-down expectations to be grossly wrong. According to this theory, the degree of expected uncertainty is signaled by the amount of one neurotransmitter (acetylcholine) that causes us to pay less attention to our models—our preconceptions—and more to what we are actually experiencing. The neurotransmitter that signals unexpected uncertainty (norepinephrine), on the other hand, has the additional effect of causing us to shift attention in order to search for new cues that might be predictive in the new context, thus helping us to form new models rather than just tune the old ones. Increased levels of norepinephrine are also associated with increased anxiety.

If something like this theory is correct, then what is responsible for the anxiety we are feeling is not just uncertainty; it is unexpected uncertainty. Maybe a measure of comfort can be taken from the possibility that the brain machinery causing our anxiety is also encouraging us to look beyond the information streams we have been relying on up to now as we try to form new models that can guide us as we move forward. I would put more trust in computational decision aids if they included this same kind of machinery.

NEW FACULTY – – – – – – – – *continued from page 1*

"I am absolutely delighted to welcome Andrew, Yannis, David, and Michael as they join our faculty," says Department Chair Andrew Barto. "No department can thrive without continuing to add new faculty, and the addition of these outstanding scholars will both deepen and broaden the department's strengths."

Andrew McGregor will join the department as an Assistant Professor in January. His research area is algorithms and complexity with applications to databases and data mining, networking monitoring, and learning. More details on Professor McGregor's research as well as his background will be highlighted in our upcoming spring issue.

Yannis Smaragdakis joined the department as an Associate Professor this Fall. His research interests are in applied programming languages and software engineering.

"My research aims to make software more modular, more reusable, more reliable, and more efficient," says Smaragdakis.

Much of his work deals with domain-specific languages and compilers, program generation and transformation, and language support for software modules. All these are closely related mechanisms, as they have in common the need to separate and factor out distinct aspects of program functionality with mechanisms beyond the capabilities of traditional programming languages.

His work often has a strong performance or systems-oriented flavor. Smaragdakis has worked on language tools for distributed computing systems, on programming models for software transactional memory, as well as on lower-level system topics, such as memory management algorithms, both at the operating system level, and at the processor cache level.

A large part of Smaragdakis' research work is concerned with software analysis and testing, with topics such as combining static and dynamic analyses for better testing accuracy, producing program invariants through a combination of concrete and symbolic execution, and deriving test cases both from models and from concrete programs.

Smaragdakis received a National Science Foundation CAREER award in 2003, an NSF ITR Award in 2002, and four best paper awards.

Smaragdakis is a senior member of IEEE, a member of the ACM, and an invited member of IFIP Working Group 2.11 (on domain-specific program generation). He received his Ph.D. and M.S. in Computer Sciences from the University of Texas at Austin in 1995 and 1999 respectively, and a B.Sc. in Computer Science from the University of Crete in1993. He

continued on page 3

was most recently an Associate Professor in the Department of Computer and Information Science at the University of Oregon, and prior to that was an Assistant Professor at the Georgia Institute of Technology.

"I'm delighted to be at UMass Amherst," says Smaragdakis. "The department has a long tradition of excellence in computing research and a superb faculty culture. I'm looking forward to lots of fruitful collaborations!"

David Smith joined the department in September as a Research Assistant Professor. He will be working with the Center for Intelligent Information Retrieval. His research interests include machine translation, natural language parsing, semi-supervised machine learning methods, and digital libraries.

"I design algorithms for analyzing and exploiting the information latent in human language," says Smith. In addition to investigating the linguistic process for its own sake, he has applied language technologies to translate between languages and to search and organize digital libraries.

Smith studies translation in a broad sense: between natural languages, between different utterances in the same languages (e.g., questions and answers), and between different layers of language ability, such as syntax and semantics. Smith develops statistical models and efficient algorithms to capture useful aspects of the translation process. In the area of digital libraries, he has built linguistic tools to help readers browse and search foreign-language text, as well as information extraction and visualization systems.

"Our ability to speak and understand language combines competence in phonology, morphology, syntax, and semantics, to name a few. After tackling them separately for a long time, we're starting to model them together," says Smith. In the spring, he looks forward to teaching a course on machine translation.

Smith has served on the program committee of conferences in natural language processing and machine learning, such as ACL, EMNLP, and ICML. In 2006, he gave a tutorial on statistical machine translation at the meeting of the Associate for Machine Translation in the Americas. He received the NSF Graduate Research Fellowship (2003-06) and the Wolman Fellowship (2002-03).

He will receive his Ph.D. in Computer Science this fall from Johns Hopkins University. He received his A.B. summa cum laude in Classics (Greek) from Harvard University in 1994. Prior to joining UMass Amherst, Smith was in the Center for Language and Speech Processing at Johns Hopkins. Previously, he was the head programmer in the Perseus Digital Library Project, one of the largest humanities digital libraries, and also held an internship at Google in the Machine Translation group.

"UMass Amherst is so outstanding, and so deep, in information retrieval and extraction, and machine learning generally," says Smith. "I'm already enjoying collaborations with faculty and students on several projects that pull together research interests that could only have been combined here."

Michael Zink, who joined the department in 2004 as a Postdoctoral Fellow and later as a Senior Research Scientist, began this fall as a Research Assistant Professor. His research interests cut across a wide range of topics in distributed systems, including the fields of sense-and-response sensor networks, distribution of high-bandwidth, high-volume data, network measurements and the design and analysis of long-distance wireless networks.

Further research interests are in wide-area multimedia distribution for wired and wireless environments and network protocols. Zink is currently a member of the leadership team of the UMass Amherst-led NSF Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (CASA).

"As Technical Integration Leader for CASA's first four-node radar sensor network, I manage an interdisciplinary team of electrical engineers, atmospheric scientists, social scientists, and fellow computer scientists," says Zink. The team is responsible for the design, implementation, and operation of CASA's Distributed Adaptive Collaborative Sensing system, which has been operational since spring 2006.

Zink's current research on wireless networks is focused on the investigation of 802.11-based, long-distance mesh networks. In this work, he has shown the impact of directional antennas in multi-hop links on overall data throughput.

In prior work on his Ph.D. thesis, "Scalable Internet Video-on-Demand Systems," Zink co-developed the streaming platform KOMSSYS that allowed the testing of new distribution mechanisms for multimedia data in the Internet. In addition, he conducted the first subjective assessment of the influence of variations in layer-encoded video on the user's perceived quality.

Yannis Smaragdakis

David Smith

Michael Zink

"I have been working in the CS department at UMass Amherst for the last four years which gave me a wonderful experience," says Zink. "The department offers a very cooperative environment with highly motivated, world-class researchers, and very talented students. I'm looking forward to fruitful collaborations with other faculty in the department and the possibility of advising and educating our students."

Zink received his M.S. (Diploma) in Electrical Engineering and Information Technology and his Ph.D (Dr.-Ing) in Computer Science from Darmstadt University of Technology in 1997 and 2003 respectively. Previously he was a researcher at the Multimedia Communications Laboratory at Darmstadt University of Technology. In a prior position, Zink was a Guest Researcher at the U.S. National Institute of Standards and Technology.

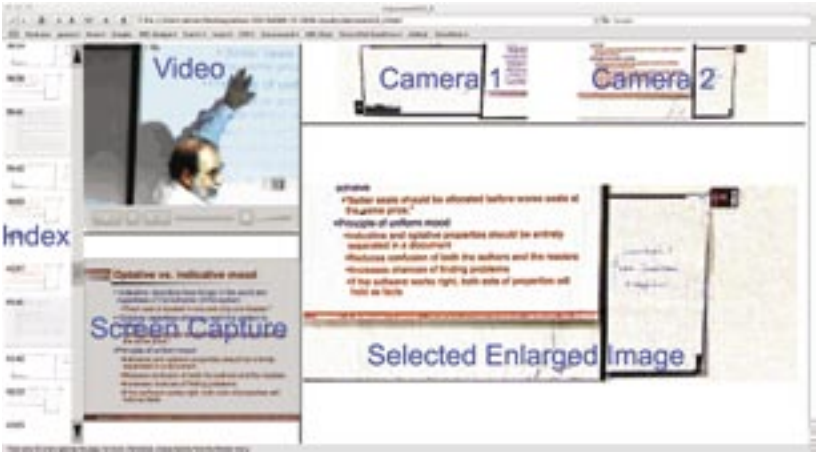
ADRION – – – – – – – *continued from page 1*

As noted, MANIC initially was the "driving application" for networked multimedia information systems research led by Kurose and Ph.D. student Jitendra Padhye (now at Microsoft). When Adrion took over day-to-day management of the MANIC Project in 1996 while Kurose was on sabbatical, the group began to look more deeply at how MANIC might be used to further teaching and learning. Research Professor Beverly Woolf and her student Mia Stern (now at IBM) were looking at how to incorporate intelligent tutoring techniques to sequence and tailor MANIC presentations by modeling student learning. Burleson and his students were investigating how to incorporate assessment in multi-media courseware. Adrion was using MANIC-delivered lectures to free up time for problem-focused learning in the classroom. PEEAS was increasing its use of MANIC as an alternative to videotape and moving toward DVD delivery of distance education. ACSIOM (a technology transfer company founded by the department) was developing MANIC presentations for the Department of Defense, museums, the arts, and outreach.

As the demand for delivering content increased, more attention was paid to supporting content authoring, adds Adrion. Finally, the then limits on network bandwidth and platform/plugin diversity required alternative solutions and an internet-enabled, CD-based version was developed. Key contributors to the MANIC projects included software engineers Ken Watts (now at TEKsystems) and Agustín Schapira (CS KDL) and graduate students Kim DeVries (Cal State), Cris Pedregal Martin (Oracle) and Santhosh R. Thampuran (Bloomberg L.P.).

By 1999, AML became RIPPLES, and the Project entered a phase of extensive production of CD-MANIC and Web-DVD courseware led by Wendy Cooper, RIPPLES manager, and Ken Watts. In the early days, MANIC was used to deliver a limited set of course materials to more than 25,000 users in more than 100 countries. During this production phase, MANIC technologies were used for over 30 full-semester courses, a number of short tutorial modules, and modular courseware for ECE, CS, Statistics, Nutrition (in cooperation with the OWL Project), Neuroscience, Entomology, Art History, and English at UMass Amherst, Smith College, NC State University, Polytechnic University and at other institutions. The RIPPLES group also began to develop multimedia presentations for the NSF Collaborative Adaptive Sensing of the Atmosphere Engineering Research Center, the NSA Center of Excellence Information Assurance Lecture Series, the Sidney Topol Distinguished Lecturer Series, the Citizen Planner Training Collaborative, and many others.

Extensions to support constructivist pedagogies made it possible to apply RIPPLES technologies in traditional and constructivist contexts. RIPPLES developed an extended search mechanism that in addition to searching within the MANIC-delivered content compiles a query based on the relevant content and returns Google search results. Notation systems provide for asynchronous and synchronous personal and collaborative note taking. Users can modify a view, for example, by selecting a specific source (video, laptop, whiteboard) and/or resizing the images. In addition to graduate student research, RIPPLES was a training ground for more than 20 undergraduates, many of whom went on to graduate studies at such institutions as Stanford, MIT, Tufts, Fordham, and UMass. Key contributors included Ken Watts, Dula Kumela (Microsoft), Esha Ray (Sun Microsystems), and Byron Wallace (Tufts), who led the jMANIC development while an undergraduate and who began as a 10th-grade student intern with RIPPLES.



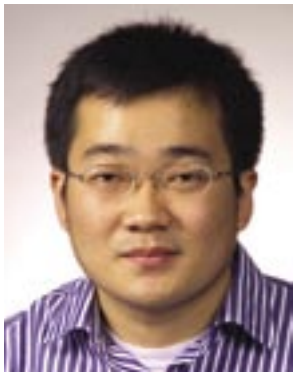
FlashMANIC delivering PAOL content

"Through much of the history of AML/RIPPLES, human operators captured audio/video content, while student, faculty, and staff content creators/editors developed text and graphic content," says Adrion. "With the demise of the video instruction program, RIPPLES has recently developed an automated content capture system." The image above is of the current FlashMANIC content delivery system with content supplied by the Presentations Automatically Organized from Lectures (PAOL) capture system.

PAOL is unique—no other system can capture unconstrained computer-based materials, capture whiteboard writing and drawing, and produce a speaker video without requiring preinstalled software, electronic whiteboards, or special training. PAOL uses high-resolution cameras, a computer screen capture device, and a wireless microphone to automatically create multimedia Flash presentations, which include a digitally edited instructor video, enhanced images of all material presented by computer or written/drawn on a whiteboard, and an index to support navigation. PAOL has also been used to deliver screen capture images in real time to students in class as part of the AutoPresenter project. The PAOL system has been used extensively during system development in several varied settings. This semester PAOL/FlashMANIC is being used to capture and deliver content for two CMPSCI classes: CMPSCI377 (Operating Systems) and CMPSCI520 (Software Engineering). This will allow RIPPLES to go beyond prior usability studies to determine the most effective presentation formats and interfaces and our efforts to capture a full course and evaluate the impact on teaching and learning. Paul Dickson (Hampshire College) led the PAOL project with his Ph.D. advisors, Adrion and Allen Hanson. Scott Myers (MIT) developed AutoPresenter and Bill Ryan (UMass Amherst CKC) developed FlashMANIC.

Diao and Wang receive NSF CAREER awards

Yanlei Diao



Rui Wang

Assistant Professors Yanlei Diao and Rui Wang have each received five-year National Science Foundation (NSF) Faculty Early Career Development (CAREER) program awards, the NSF's most prestigious award for new faculty members. RFID (Radio Frequency Identification) research and improved computer graphics research are the focus of their CAREER grants.

With her 2008 CAREER award, "Efficient, Robust RFID Stream Processing for Tracking and Monitoring," Diao is working on designing and developing an efficient, robust RFID stream processing system that addresses the challenges in emerging RFID deployments, including the data-information mismatch, incomplete and noisy data, and high data volume. Such systems will enable large-scale, real-time tracking and monitoring in supply chain management, healthcare, pharmaceuticals, and library management.

Diao's project has two main contributions. To handle high-volume, incomplete, noisy data, she first develops a low-level inference and compression substrate over RFID streams. This substrate infers locations of unobserved objects and inter-object relationships from incomplete, noisy raw data using probabilistic algorithms. To handle high data volume, it performs online inference, enabling online compression by identifying and discarding redundant data. The second contribution is higher-level complex event processing that addresses the data-information mismatch by encoding application information needs as complex event patterns and evaluating these patterns continuously over event streams. The project offers a foundation for complex event processing including theoretical underpinnings, automata-based mechanisms for efficient evaluation, and techniques for robust processing in the face of many data quality issues.

"The proposed research explores a set of fundamental issues at the nexus of stream query processing, theory, and machine learning, which can be generalized beyond the RFID-based concrete setting," says Diao. "Our project integrates research and education through curriculum development and teaching and research lab development. It also enables broader participation of women and minorities in research through college outreach and CRA-W's distributed mentor program."

Diao joined UMass Amherst in 2005 after receiving her Ph.D. in Computer Science from the University of California, Berkeley. Her research interests include information architectures and database management systems, with a focus on data streams, sensor data management, data dissemination, XML query processing, and learning-based data processing.

Assistant Professor Rui Wang received a CAREER award, "Nonlinear Processing of Light Transport Data for Realistic Computer Imagery," to support his research on improved computer graphics.

"Realistic computer graphics involving very large image datasets—such as lightfields, precomputed light transport, time-lapse video, and measured material database—has become an increasingly important and interesting topic in recent years," says Wang. "While rendering systems can exploit these data to achieve greatly improved performance and visual fidelity, such datasets also present significant challenges for efficient processing, not only because of their large scale and high dimensionality, but more importantly, because of their intrinsic nonlinear structures. That is, the data naturally contain very high-frequency, rapidly changing information that cannot be adequately represented using linear processing models that prevail elsewhere in computer graphics."

Wang is addressing these issues in his CAREER project that deals with computer graphics research to develop nonlinear processing methods for large-scale, multidimensional light transport data. These nonlinear methods adaptively select the best solutions for representation driven by the data's intrinsic nonlinear structure, thereby significantly improving the sampling, compression, reduction, and reconstruction of such data.

Wang's project involves three steps of research. The first step collects light transport data from both simulated and measured sources and derives a novel sampling strategy by using data-driven analysis. In the second step, he investigates several new methods for nonlinearly reducing and approximating light transport data into compact representations. The final phase develops efficient computational methods using the compressed data representation for fast reconstruction and real-time visualization. The resulting new approaches from this research can leverage broader applications in other fields that involve similar types of image datasets, such as geosciences, medical imaging, and forensic analysis.

Wang joined UMass Amherst in 2006 after completing his Ph.D. in Computer Science at the University of Virginia. His research interests are in the field of computer graphics, with focuses on global illumination algorithms, interactive photorealistic rendering, and appearance modeling.

The CAREER program recognizes and supports the early career-development activities of those teacher-scholars who are most likely to become the academic leaders of the 21st century. Previous department faculty CAREER award recipients include Micah Adler (2002), Emery Berger (2004), Oliver Brock (2006), Mark Corner (2005), Deepak Ganesan (2006), Erik Learned-Miller (2006), Brian Levine (2001), Sridhar Mahadevan (1995; awarded at Univ. of S. Florida), Kathryn McKinley (1996; now at UT-Austin), Gerome Miklau (2007), Prashant Shenoy (2003), Ramesh Sitaraman (1997), and Shlomo Zilberstein (1996).

SIGCOMM honors Towsley for lifetime achievement

Distinguished Professor Don Towsley is being recognized for his contributions to the field of computer networking by the Association of Computing Machinery's special interest group on data communications (ACM SIGCOMM).

Established in 1989, the SIGCOMM Award is awarded annually to a person whose career efforts represent a significant contribution to the field and a substantial influence on the work and perceptions of others in the field. Towsley is being honored for contributions to the modeling, analysis and control of communication networks.

The award, which consists of an engraved pyramid from Tiffany's and an honorarium, was presented Aug. 19 during ACM SIGCOMM's annual conference in Seattle. As the honoree, Towsley gave a keynote address at the meeting.

Over a research career spanning three decades, Towsley has made innovative and pioneering contributions in developing foundational modeling and analysis techniques that have enabled a better understanding of some of the most important aspects of today's computer networks, network protocols and networked applications. His research contributions, embodied in part by a collection of nearly 200 journal papers that he has authored, exhibit great diversity across scientific disciplines and include a long list of collaborators. Highlights of his work include the first empirically-validated analytical model of the TCP protocol, foundational work on network tomography to infer with statistical rigor internal network behavior based on end-to-end observations, the use of fluid models and stochastic differential equations to elucidate network performance in the presence of competing sessions, methods to smooth and streamline delivery of multimedia content across the Internet, and key contributions to advance the state of the art in reliable multicast protocol design.

Towsley is also known for his dedication and longstanding service to the community. In addition to the dozens of students he has mentored, many of whom themselves have made lasting contributions to the field, he has also organized numerous workshops and symposia, served on editorial boards for top journals and countless technical program committees, and currently serves as editor-in-chief of the IEEE/ACM *Transactions on Networking* journal.

Towsley is a fellow of the ACM and the Institute of Electrical and Electronics Engineers (IEEE). Last year, he received the IEEE Koji Kobayashi Computers and Communications Award and the ACM Special Interest Group on Measurement and Evaluation (SIGMETRICS) Achievement Award, both given for lifetime research achievements. Towsley received the 1999 IEEE Communications Society William Bennett Award and several conference and workshop best paper awards. He is also the recipient of the University of Massachusetts Amherst Chancellor's Medal in 2001-2002, and the College of Natural Science and Mathematics Outstanding Research Award in 2003.

This fall, Towsley received two additional campus awards. He was among seven nationally known faculty who received the Award for Outstanding Accomplishments in Research and Creative Activity during the UMass Amherst Faculty Convocation, and he received a UMass Amherst 2008-2009 Samuel F. Conti Faculty Fellowship Award. As part of the Conti Award, Towsley received a year's leave of absence to concentrate on activities related to graduate education, research, creative work, and scholarly attainment.

Test of Time Award

UMass Amherst CS Ph.D. alums Jitendra Padhye ('00) and Victor Firoiu ('98) and Distinguished Professors Don Towsley and Jim Kurose received the 2008 ACM SIGCOMM Test of Time Award for their paper "Modeling TCP throughput: a simple model and its empirical validation," which was presented at the 1998 ACM Special Interest Group on Data Communication (SIGCOMM) Conference.

According to ACM, the ACM SIGCOMM Test of Time Award recognizes a paper published 10 to 12 years in the past in ACM *Computer Communication Review* or any SIGCOMM-sponsored or co-sponsored conference that is deemed to be an outstanding paper whose contents are still a vibrant and useful contribution today.

NSM award for Lesser

At the fall convocation of the UMass Amherst College of Natural Science and Mathematics (NSM), Professor Victor Lesser received the Outstanding Faculty Award in the Area of Research for his contributions to artificial intelligence in the areas of problem-solving architectures, multi-agent systems, real-time AI and signal understanding.

Lesser, the director of the Multi-Agent Systems Lab (MAS) and member of the faculty since 1977, was also among seven nationally known faculty who received the Award for Outstanding Accomplishments in Research and Creative Activity during the UMass Amherst Faculty Convocation held this fall.

Kurose named Interim Dean of NSM

Photo by Ben Barnhart

Distinguished Professor Jim Kurose has been named Interim Dean of the College of Natural Sciences and Mathematics (NSM) for a two year term.

Kurose co-directs the Networking Research Laboratory and is associate director of the National Science Foundation-supported Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere. His research interests include computer network protocols and architecture, network measurement, sensor networks, multimedia communication, and network modeling and performance evaluation.

He is the recipient of the Taylor Booth Education Medal by the Institute for Electrical and Electronic Engineers (IEEE). Kurose is also an eight-time recipient of the Outstanding Teacher Award from the National Technological University and has also received teaching awards from the Northeast Association of Graduate Schools and the College of Natural Sciences and Mathematics. This past summer, he taught at the Summer Leadership Institute in Mysore, India.

Kurose is an elected fellow of the IEEE and the Association for Computing Machinery, and he is currently serving a three-year term on the Computing Research Association Board of Directors.

Appointed a Distinguished Professor in 2004, Kurose joined the faculty in 1984 after receiving his Ph.D. at Columbia University. He chaired the Computer Science Department from 1998 to 2001.

Renovations almost complete

Did you know that the department still occupies a portion of its old site, the Lederle Graduate Research Center (LGRC), in addition to the "new building," the Computer Science Research Center? The department occupies much of the third floor of the LGRC "lowrise," in addition to space in the "tower" which houses the CS Educational Laboratory (EDLAB), recently upgraded with the addition of 30 new triple-bootable IMACs.

Renovation of the third-floor LGRC space has been in the works for several years and is almost complete. The renovations will bring the laboratory and office space up to the standards of the space in the Computer Science Research Center.

The third-floor LGRC space houses research labs, including the Laboratory for Perceptual Robotics, a part of the Computer Vision Research Lab, graduate and undergraduate student office and study space, faculty and staff offices, and classroom/conference space. The Center for Educational Software Development (CESD) occupies one corner of the third floor, and the Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) occupies another.

Renovations include refurbishment of rooms with new paint, carpeting, ceiling tiles, light fixtures and emergency lights. In addition, electrical and air conditioning systems in the LGRC server room were upgraded to accommodate "Swarm," the department's new high performance cluster with 60 processors/480 cores and 65 TB of disk space.

Included on the third floor is a space that the department intends to turn into a modern flexible study/classroom space to be used for small classes and seminars. The department hopes to furnish the room with 15–20 comfortable and versatile chairs to give the room the maximum flexibility while maintaining a modern look.

The renovations have been ongoing for two years. While he was department Chair, Distinguished Professor Bruce Croft worked with the campus administrators on the initial planning and funding for the renovations, and current Chair Andrew Barto is working towards completion of the project.

While the renovations were funded by the campus, there is still a need for funds to furnish the rooms. Your donation can help to get the renovations completed so that our students can fully use this space. During these tough financial times, we really need your help to complete these projects that will benefit generations of CS students.

Influential paper award

Professor Eliot Moss, along with Brown University co-author Maurice Herlihy, received an award for the most influential paper at the International Symposium on Computer Architecture (ISCA) held in Beijing this past June.

Moss and Herlihy were honored by the ACM Special Interest Group on Computer Architecture (SIGARCH) and Institute of Electrical and Electronics Engineers-Computer Society Technical Committee on Computer Architecture (TCCA) with the ISCA Influential Paper Award for their 1993 paper, "Transactional Memory: Architectural Support for Lock-Free Data Structures."

Moss's current research is further extending the flexibility and concurrency of transactional memory, looking at *nested transactions,* and in particular at *open nestings*. This allows short transactions to be composed into larger ones while managing their conflicts and their rollback at a logical rather than physical level. Open nesting is complex and depends on correct implementation of transactional parts of applications, but it greatly increases possible concurrency.

According to SIGARCH, the Award recognizes the paper from the ISCA Proceedings 15 years earlier that has had the most impact on the field (in terms of research, development, products, or ideas) during the intervening years.

Professor Paul Utgoff (1951 – 2008)

The department is very saddened to report that Professor Paul E. Utgoff died on October 11, 2008 at the age of 57 due to complications from surgery that he elected to undergo as part of his battle against appendiceal cancer.

Professor Utgoff joined the department as an Assistant Professor in the fall of 1985. He was promoted to Associate Professor in 1991 and to full Professor in September 2008.

"Paul was a genuine scholar who never lost sight of the deepest problems in his field, and he was an essential force in helping the department make important decisions. I cannot express our sadness at this loss of a truly valued colleague and friend," said Department Chair Andrew Barto.

Utgoff is recognized internationally as a pioneer and leader in the area of machine learning. Among the topics he addressed is the longstanding fundamental problem of how intelligent systems can acquire features, terms, and other representational structures that are prerequisite to further learning. Utgoff started his work in this area as a Ph.D. student at Rutgers where he developed a system that could modify its own representation language by adding terms to repair faulty generalizations, an accomplishment now recognized as a seminal contribution to the field. His subsequent research demonstrated his skill in selecting key problems and his persistence in pursuing elegant solutions. Three areas of contribution are particularly notable for their impact on the field: (1) his work on incremental decision tree induction (2) multivariate decision trees, and (3) problem representation.

In 1993, Utgoff designed the Incremental Tree Inducer (ITI) program and a new algorithm called Direct-Metric Tree Inducer (DMTI) made feasible by the ability to restructure trees efficiently. He made his code available on the Web in October 1994, and maintained it ever since. It has been downloaded 6,530 times. His work remains the state of the art in incremental decision tree induction. Returning to his interest in problem representation, his recent research focused on "many-layered learning" following the principle that all learning is simple if the prerequisites are in place.

Utgoff recently wrote, "…the fascination I have had in representation induction since my doctoral dissertation remains at the heart of my research interests. The great majority of my work has been motivated by … my desire to see machines learn difficult concepts and skills in a sustained manner. … I have been working on representation repair and modification, term construction, feature construction of various kinds, goal regression to identify essential sets, learning deeply layered knowledge from data or from textbooks, learning from expert choices, hybrid representations, and incremental decision tree induction."

In other recent work, Utgoff was a co-PI on the NSF ITR collaborative project "Interactive Software Systems for Expert-Assisted Image Analysis and Classification of Aquatic Particles." In this project, he worked with the Computer Vision group and Bigelow Laboratories building systems to perform various classifications of phytoplankton and zooplankton from digital images obtained from ocean studies.

Utgoff served as an action editor for the *Machine Learning* journal from 1991-1995, with membership on the editorial board through 1997. He also has served in many organizational roles for the International Conference on Machine Learning (ICML). He chaired and hosted the Tenth ICML at UMass Amherst, and served as an Area Chair for the Conference in 2000 and 2002. Most recently, he was the area chair of the 2007 European Conference on Machine Learning. In addition, he has held many leadership positions on university and departmental committees, including chairing the University Research Council from 2005 to 2007 and his essential contributions to the overhaul of the undergraduate curriculum that the department expects to unveil in the Spring 2009 semester.

Professor Utgoff earned a Bachelor of Music in pipe organ performance from the Oberlin College Conservatory of Music in 1979 and a Ph.D. in Computer Science from Rutgers, The State University of New Jersey, in 1984. Following two years as a Research Scientist at Siemens Research and Technology Laboratories, he began his career at UMass Amherst. His lifelong passion for music manifested itself in recent work on how artificial intelligence meets various topics in music, such as music perception, cognition, and understanding, as well as his work as Music Director at Saint Philip's Episcopal Church in Easthampton. He was an avid backgammon player and as a young man piloted light airplanes, qualifying as an instrument flight instructor.

He is survived by his wife Karen and children Naomi Utgoff, Emily VanHassel, and Ariel Utgoff.

A service celebrating Professor Utgoff's life was held on November 2nd in Saint Philip's Episcopal Church, Easthampton. Donations may be made to the Cooley Dickinson Hospital in Northampton; the UMass Amherst Department of Computer Science, where funds will be used for scholarships to support promising students; the Memorial Fund at St. Philip's Episcopal Church in Easthampton; or a charity of the donor's choosing.

CAITE reaches out to students and educators

Over a dozen girls from the Holyoke chapter of Girls Inc. visited UMass Amherst on August 8, 2008 to tour information technology (IT) workplaces and learn about computer-related careers. The Holyoke chapter of Girls Inc. participates in a national program called "Build IT" that is designed to tap into the curiosity of girls between the ages of nine and 15 about design and communication technology.

The visit was arranged by the Commonwealth Alliance for Information Technology (CAITE), led by Professor Rick Adrion, along with the academic computing department at the Office of Information Technology. Girls Inc. is a non-profit organization dedicated to inspiring girls to seek education and meet physical and intellectual challenges.

The girls received details on a variety of computing careers including instructional design and faculty support, video editing and digital photography, media lab management and Web communications management. The girls also met with Research Professor Beverly Woolf and Evelyn Barney in the Center for Knowledge Communications lab. They saw a demonstration of the lab's intelligent tutor project, where computerized tutors respond to students and adjust to their emotional state by using an array of sensors.

As part of its effort to diversify computing, CAITE and the Commonwealth Information Technology Initiative (CITI) also hosted two professional development workshops for educators this July in collaboration with the NSF BPC Alliance Georgia Computes!. Both workshops were led by Mark Guzdial, Associate Professor in the College of Computing at Georgia Institute of Technology and PI of Georgia Computes!. Educators learned alternatives to teaching introductory computing courses to college and high school students

Iris Chelaru of OIT talks with participants about communications tools.

Twenty-three faculty and teachers participated in a three-day workshop at Microsoft's facility in Waltham, MA. "Introduction to the Media Computation Approach" explored a fresh approach to teaching introductory computing in Python, Java, and data structures using Guzdial's dynamic Media Computation approach to teaching.

Later that week, Guzdal led a two-day workshop titled "Innovative Approaches to First Courses in Computing" at the UMass Amherst Computer Science building. The workshop presented five different approaches to introducing computing to students, all approaches that focus on a particular context and have been successful at attracting and retaining women and minority students.

To promote information technology education, CAITE has launched a Facebook site which provides updated information on services and events such as those described above.

Weems develops school curriculum

Associate Professor Chip Weems is developing a computer science curriculum for Hartsbrook High School in Hadley, MA which will eventually be presented for wider use in Waldorf high schools around the country.

"Because Waldorf education is a developmentally-based pedagogy, I've been working on a series of short courses, one at each of the four grade levels," says Weems. "For example, at the 9th grade level, the focus is concretely on what the computer is, so we look at digitization, programmability, and limits of computation. In contrast, at the senior level, we explore more abstract, philosophical, and societal aspects of computing."

Because these courses are required, Weems' goal is to foster the kind of understanding that pierces the mystery and hyperbole that often surround computing in the media, as well as to generate interest in further study of computer science.

There are over 50 Waldorf high schools in the United States, but although they have common curricular elements in other subjects, computer science is not taught with any consistency. Once the Hartsbrook curriculum has been fully developed and tested, Weems plans to make it available for broader use. "There is potentially a significant impact from this effort because surveys have found that 43 percent of Waldorf high school graduates go into the sciences, as compared with an average of 18 percent from public schools, and of those, a large fraction go on to graduate school," adds Weems.

Research program for undergrads

This summer, the department, under the direction of Professor Rick Adrion, offered a 10-week National Science Foundation summer Research Experiences for Undergraduates (REU) program. Eleven students, from schools such as Wellesley College and the University of Puerto Rico Mayaguez, spent the summer on campus doing research in many areas of computer science. Pictured are the REU students along with the students' mentors.

Cisco Systems, Inc.

Cisco is actively involved in supporting the department's research. Recently, the company provided research funding to Assistant Professor Yanlei Diao for her work on "In-Network Event Processing over Distributed Streams," and it also funded Associate Professor Brian Levine's project on "Municipal Wireless as a Forensics Sensor."

In addition, Cisco awarded two scholarships to CS students this fall. Alison Bird and Tiffany Chao received Cisco Systems Undergraduate Scholarships targeted towards underrepresented groups in computer science. Cisco also sponsors the ACM student chapter meetings and was instrumental in getting the new chapter established in 2008.

EMC^2 Corporation

In September, EMC^2 spent a day at UMass Amherst at various departments on campus. In the Computer Science Department, Daniel Bailey gave a research talk on "Wireless Authentication Token Protocols." Bailey is a Senior Research Scientist at RSA Laboratories, part of the Security Division of EMC^2. UMass Amherst alum, James Pearson, EMC^2 Vice President introduced Bailey and spoke briefly about the current state of EMC^2.

In addition to bringing technical speakers to campus, EMC^2's RSA Labs continues to support the RFID Consortium on Security and Privacy both academically and financially. Led by Assistant Professor Kevin Fu, RFID CUSP is a partnership between academic and industrial scientists specializing in RFID security and privacy. Its mission is to make RFID safe for consumers by conducting open research and educating the next generation work force that will develop, deploy, and maintain secure RFID infrastructures.

Cisco Scholarship Recipients

Computer Science undergraduates Tiffany Chao and Alison Bird (l. to r.) received Cisco Systems Undergraduate Scholarships targeted towards underrepresented groups in computer science.

Google

In October, Google Research Scientist Gideon Mann gave a talk on "A Baseline for Large Vocabulary Video Annotation." Mann is a former Post-Doctoral Researcher in the department's Information Extraction and Synthesis Laboratory, and he is still in active collaboration with that group. In addition to the technical talk, Google hosted a reception in which other Googlers, including CS alum Vitaliy Lvin (MS '07), were available to answer questions about the company and its research.

McKesson Corporation

McKesson Corp., the department's newest industrial affiliate, visited in October to give students an overview of the company, located just down the road in Hadley, MA. During an afternoon seminar, McKesson researchers discussed the current research being undertaken at the world's largest health care services company. Their office started in the 70's as Amherst Associates, a spin-off company from the UMass Amherst School of Public Health. Their software caught the attention first of HBO & Company and then McKesson who acquired them in 1998. They are in the Performance Management/Business Intelligence domain. McKesson visitors included Steve Yarrows, and CS alums Sally Waisbrot (M.S. '81) and Kevin Peret (B.S. '08). Waisbrot, McKesson's Director of Research and Development, also spoke at a departmental seminar in the spring.

Outstanding Synthesis Projects

Dov Katz and Marek Petrik received the 2007-2008 Awards for an Outstanding Synthesis Project, sponsored by Yahoo!. Katz's project was "Manipulating articulated objects with interactive perception," and Petrik's project was "A linear programming approach to bounds and basis construction for Markov decision processes."

Yahoo! Inc.

Yahoo! is partnering with Professor James Allan on BOSS (Build your Own Search Service). Yahoo!'s open search web services platform gives researchers the ability to conduct open research on search engines that was impossible in the academic environment. The department was one of seven universities initially chosen to partner with Yahoo! on the BOSS project.

In addition to sponsoring the department's weekly Machine Learning & Friends technical seminar series luncheons, Yahoo! researchers visited the department to present their work. CS alums Fernando Diaz (Ph.D '08) and Hema Raghavan (Ph.D. '07) spoke in the department this fall. Both are Research Scientists at Yahoo!

Yahoo! also actively supports the department's students and activities. The company provided funding for the CS Women's group, the ACM chapter, travel funds for the Grace Hopper Conference, and they also sponsor the department's outstanding undergraduate student awards and graduate student thesis and synthesis project awards. In the spring, Yahoo! sponsored a foosball tournament to make use of the foosball table they donated to the CS department.

Emotional intelligence for computerized tutors

Computer Science researchers are developing interactive computerized tutors that sense a student's emotional and motivational state of mind, while presenting information designed to appeal to the student's intellectual curiosity. Special sensors are used to help the computer tutor respond when students become angry, frustrated, or bored, based on body language, attention, and other indicators.

Research Professor Beverly Woolf says the introduction of the emotion sensors helps the scientists respond to how people actually learn. "Emotion and cognitive functions are strongly correlated," Woolf says. "So if you improve the social intelligence of the computer, students respond the way they would to another person. Sensors allow the computer to identify students who pay attention and those too tired or bored to learn. Using these cues, the computer provides individualized instruction."

Woolf is part of a team that includes Research Scientist Ivon Arroyo and Professor Andrew Barto, and Winslow Burleson from Arizona State University. The tutors they have designed teach high school students geometry and algebra, but can be adapted to any subject, Woolf says. The work is funded by grants from the National Science Foundation.

Woolf says the non-invasive sensors replicate what top-notch human teachers do in the classroom to engage their students. "Master teachers devote as much time working on a student's motivation as they do on straight teaching," she says. "They understand that students who feel anxious or depressed don't assimilate information properly."

The sensors they are developing include a camera that views facial expressions. Woolf says certain expressions on students' faces or how they tilt or hold their head are strong indicators of their level of interest in what they are doing. There is also a posture-sensing device in the seat of the chair that measures the amount of fidgeting, or stillness, which are other indicators of interest and concentration on the task.

There is also a pressure-sensitive computer mouse that can tell how hard the user is pushing down. Previous research has shown that users who find a computer task frustrating often apply significantly more pressure to the mouse than those who do not find the same task frustrating, Woolf says.

In addition, a wireless skin conductance wristband worn by the student shows the level of arousal. A certain amount of arousal is a motivator toward learning and tends to accompany significant, new, or attention-getting events, she says.

The students, equipped with the emotion sensors, are presented mathematical subject matter for a period of up to two hours, Woolf says. During each session, the computer analyzes the information it gets from the sensors and adjusts how it presents the subject matter. Sometimes, that means halting the program and offering the student an alternative activity to reignite interests. Alternatively, the computer may go back and revisit material that the student has failed to master.

Sensors can also detect when a student tries to "game the program" by randomly choosing answers or hurrying through the problems. When such behavior is detected, the computer tutor responds in a friendly manner by asking the student to slow down or read more carefully.

Undergrad Profile: Tommy Boucher

Tommy Boucher first became interested in computer science in high school at the Massachusetts Academy of Mathematics and Science. He later worked professionally as a software developer.

Boucher spent five years in the Marines, attained the rank of sergeant, and completed two full tours of duty in Iraq, in Ramadi, Fallujah, and Baghdad. He led a team of developers creating programs from concept to deployment, including a personnel management system, a production tracking system, and a data warehouse system that stored every intelligence report created by the Marine Corps for Operation Iraqi Freedom.

After this, Boucher came to UMass Amherst as a Computer Science undergraduate. Last year, he began working with Professor Jack Wileden to start a UMass Amherst ACM student chapter and was elected as the club's first chair. Daniel Amirault was vice chair, Ali Shah was treasurer, and Daniel Roy was secretary. The club received funding from Cisco Systems for ACM memberships and meetings. "We built a strong student base and became a nationally recognized chapter of the ACM," says Boucher. The AY '08-'09 ACM chapter leaders are chair Daniel Roy, vice-chair Daniel Amirault, treasurer Leon Moreyn, and secretary Stevie Sellers.

In addition to his coursework, Boucher works as a software developer in Northampton. He still finds time to bicycle, hike, and play jazz. He plays trumpet and piano and is trying to learn the saxophone. After graduating, he plans to go to graduate school to receive a Ph.D. in mathematics.

Second place in ACM programming contest

Undergraduates Eric Price, Dan Amirault, and Dan Strom (l. to r.) placed second in the preliminary regional ACM programming contest and qualified for the regionals in Rochester, NY.

Faculty News




The UMass Board of Trustees approved the promotion of **James Allan** and **Paul Utgoff** to full professor and approved tenure and promotion to Associate Professor for **Emery Berger** and **Oliver Brock (above, l. to r.).**


Assistant Professor **Bill Verts** was promoted to Senior Lecturer II, and CS alum **John Ridgway** (Ph.D. '04) joined the department as a Lecturer for AY '08-'09 (l. to r.).

Professor **Eliot Moss** and co-authors of an OOPSLA '06 paper were honored by having the paper chosen to appear in *Communications of the ACM* (CACM) Research Highlights, part of a new "best of the conferences" series that CACM is inaugurating. Highlights of their paper, "The DaCapo Benchmarks: Java Benchmarking Development and Analysis," appeared in the September 2008 issue of CACM, the leading monthly print and online magazine for the computing and information technology fields.

Assistant Professor **Deepak Ganesan** has received a 2008 IBM Faculty Award for his research project "RiverNet: Sensing, Networking and Modeling of Data from a River Sensor Network." According to IBM, to qualify for this internationally competitive award, the faculty nominee must have an outstanding reputation for contributions in their field or, in the case of junior faculty, show unusual promise.

uBot-5 meets Bill Gates

Graduate student Patrick Deegan (right) and the Laboratory for Perceptual Robotics' uBot-5 take center stage at the Microsoft Developers Conference with Bill Gates during his last public presentation as Microsoft's Chairman.

Professor **Shlomo Zilberstein** was selected to be the next Editor-in-Chief of the *Journal of AI Research* (JAIR), which is considered the premier journal in the field. Zilberstein was also voted the new President Elect of the International Conference on Automated Planning and Scheduling (ICAPS) Executive Council. ICAPS is an international organization whose primary goal is to maintain the ICAPS conference as the premier annual forum for researchers and practitioners in automated planning and scheduling.

CS Researchers received the Best Paper Award at the 29th IEEE Symposium on Security & Privacy for their paper "Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses." Assistant Professor **Kevin Fu**, **Ben Ransford, Shane S. Clark,** and **Benessa Defend** accepted the award in Berkeley, CA, in May. CS Alum **Thomas S. Heydt-Benjamin** (MS '08) and undergraduate **William Morgan** were also co-authors.

Distinguished Professor **Bruce Croft** gave a keynote address entitled "Unsolved Problems in Search" at the 2008 ACM Conference on Information and Knowledge Management (CIKM) in October. In addition, Croft gave a speech, "Longer Queries, Better Answers?," as part of the Distinguished Lecture Series at the Microsoft Research Silicon Valley Campus in June.

Associate Professor **Sridhar Mahadevan** published a new book, "Representation Discovery using Harmonic Analysis," which is a volume in the *Synthesis Lectures on Artificial Intelligence and Machine Learning* (edited by Ron Brachman and Tom Dietterich), Morgan Claypool Publishers.

Professors **Brian Levine** (PI), **Yanlei Diao**, and **Jim Kurose** received a Mellon Foundation Mutual Mentoring Team Grant to provide new team mentoring efforts within the department.

Associate Professor **Yannis Smaragdakis** gave a keynote address at the 15th International SPIN Workshop on Model-Checking of Software in Los Angeles this summer.

Associate Professor **Hava Siegelmann** became associate editor of *Frontiers in Computational Neuroscience*.

Assistant Professor **Erik Learned-Miller** organized a workshop, Faces in Real-Life Images, at the European Conference of Computer Vision held in Marseille, France, in October.

Associate Professor **Chip Weems** co-chaired a new Workshop on Large-Scale Parallel Processing at the 2008 International Parallel and Distributed Processing Symposium held in Miami.

Adjunct Professor **Lee Spector** edited a special issue of the journal *AI-EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing,* "Genetic Programming for Human-Competitive Designs." Spector and his team also received a gold medal in the 5th Annual "Humies" Awards for Human-Competitive Results Produced by Genetic and Evolutionary Computation, held at the 2008 Genetic and Evolutionary Computation Conference in Atlanta, Georgia.

Researcher News

The Laboratory for Advanced Systems Software welcomed Research Scientist **Emmanuel Cecchet.**

Dr. **Pak-Ching Lee** joined the Networks Lab as a Postdoctoral Research Associate.

Dr. **Ivon Arroyo**, of the Center for Knowledge Communications, was promoted to Research Scientist.

He Luo is a Visiting Scholar from Hefei University of Technology, China, working with the Multi-Agent Systems Laboratory.

The Center for Intelligent Information Retrieval welcomed **Xuanjing Huang** as a Visiting Scholar. She is the Vice Chair and Professor of Computer Science and Engineering at Fudan University.

New Visiting Scholars in the Laboratory for Advanced Software Engineering Research include Dr. **Junchao Xiao**, Assistant Professor at the Chinese Academy of Sciences, and **Danhua Wang**, a graduate student at Nanjing University.

Zhao Cao is a Visiting Researcher from Beijing Institute of Technology who is working with Assistant Professor **Yanlei Diao**.

Student News

This year's outstanding undergraduate award recipients are **Matthew Meehan** (Overall Achievement), **Paul Barba** (Artificial Intelligence), **James Barber** (Networks), **Shane Clark** (Security), **John Tuttle** (Software), **Timur Alperovich** (Systems), and **Bartholomew Parkis** (Theory). Each of the honorees graduated in 2008 with a B.S. in Computer Science.

Graduate student **Alan Carlin** and Professor **Shlomo Zilberstein** received the Best Paper Award from the AAMAS 2008 Workshop on Multi-Agent Sequential Decision Making in Uncertain Domains (MSDM), for their paper "Observation Compression in DEC-POMDP Policy Trees."

Graduate students **Colin Barringer, Jeffrey Johns**, and **Marek Petrik** placed second in the Tetris track at the International Reinforcement Learning Competition.

With the goal of integrating Japanese language and culture into her career, PRISMS graduate student **Benessa Defend** is studying Japanese for a year in Yokohama through a National Security Education Program David L. Boren Fellowship.

PRISMS graduate student **Ben Ransford** recently won a 2008 Eugene M. Isenberg Award, a campus-wide competition, for his demonstrated commitment to integrating science, engineering, and management. While it is rare for CS students to receive this award, Özgür Simşek won this award in 2002. In other news, Ransford and his wife Megan Sielken welcomed the birth of their son, Linus Foster Ransford, on October 15.

AAAI Workshop Award

The UMass Amherst CS team won third place for their robotic research presentation at the AAAI Workshop on Mobile Manipulation. Chris Dellin and Bill Townsend of Barrett Technologies, the robotic arm manufacturer, pose with CS graduate students Dov Katz and Jacqueline Kenney (l. to r.).

Graduate student **William (Tom) Billings** received this year's Robin Popplestone Fellowship in Robotics & Artificial Intelligence. He graduated summa cum laude in May 2008 from Vanderbilt University with a B.S. in Electrical Engineering.

Undergraduate **David Maletz** received a 2008 Luise Bronner Scholarship for his academic performance.

Graduate student **John Altidor** received a National Consortium for Graduate Degrees for Minorities in Engineering and Science (GEM) Fellowship when he joined the department this fall.

Marek Petrik and **Nilanjan Banerjee** were among 16 students who received University Graduate School Fellowship awards.

CS alum **Xing Wei** (Ph.D. '07) and IESL graduate student **Xuerui Wang** are the proud parents of Richard Bojin Wang, born on June 5.

Staff News

Brenton Annan joined the Computer Science Computing Facility (CSCF) staff as an Associate Software Specialist.

Tyler Trafford was promoted to a Software Specialist 1 with CSCF and the Networks Lab.

Shane Clarke joined the Privacy, Internetworking, Security, and Mobile Systems Laboratory as an Associate Software Engineer.

University of Massachusetts Amherst
Fall / Winter 2008

ALUM Matters

A newsletter for alums of the Department of Computer Science

Jacobs studies biological and artificial intelligence

Multidisciplinary research seems to be the rage these days; a search of the Internet with Google yields 698,000 hits for the keywords "National Science Foundation" and "multidisciplinary." University of Rochester Professor Robert Jacobs (UMass Amherst CS Ph.D. '90) has been pursuing multidisciplinary research on biological and artificial intelligence for many years. He has developed new machine learning algorithms and new computational models of human cognition in a variety of domains, including perception, motor control, and memory organization.

While at UMass Amherst, Jacobs was advised by Professor Andrew Barto. Afterwards, he served as a postdoctoral fellow in the lab of Professor Michael Jordan at the Department of Brain & Cognitive Sciences at the Massachusetts Institute of Technology. He then was a postdoctoral fellow in the lab of Professor Stephen Kosslyn at the Department of Psychology at Harvard University. He currently is a faculty member at the University of Rochester where he serves as Professor of Brain & Cognitive Sciences, of Computer Science, and of Vision Science.

His research activities have been shaped by the belief that biological and artificial intelligence should be studied in tandem. "No one discipline has all the answers regarding the nature of cognition and intelligence, though different disciplines have yielded different insights and perspectives," says Jacobs. "To obtain a complete understanding of intelligence, we'll need to work across traditional disciplinary boundaries. The payoff from this work will be both technological innovations that improve our lives and also new ways of thinking about our own minds."

One question that has always fascinated Jacobs is whether human cognition is "optimal" in some computational or mathematical sense. He has addressed this question in several research projects.

For his Ph.D. dissertation, he considered whether it is preferable to have a highly modular computational system in which different modules perform different tasks, or if it is preferable to have a single monolithic system that performs all tasks. He studied this issue through the development of a "mixtures of experts" system, consisting of multiple learning devices, such as multiple artificial neural networks. The system's learning algorithm uses a competitive learning scheme to adaptively partition the training data so that different devices learn different subsets of data items. His simulation results showed that there are significant advantages to highly modular systems in terms of learning speed. At the same time, however, there are also disadvantages because such systems require special mechanisms to integrate information to produce robust performances in novel circumstances.

This finding led him to think about how people integrate information from different information sources. A convenient area to study this issue is the domain of perception where we perceive our environments through multiple sensory modalities, such as vision, audition, and touch. In one set of experiments, Jacobs and his colleagues examined whether people combine visual and auditory information in a statistically optimal manner for the purpose of spatial localization. Previous research suggested that people exhibit "visual capture" —if visual and auditory information indicate slightly different locations for an event, people will localize the event to the location indicated by the visual modality, a phenomenon known as the "ventriloquist effect." This outcome reflects the fact that people are more accurate at making spatial judgments based on visual information than on auditory information. What would happen if people were placed in a virtual-reality environment in which

A subject using a visual-haptic virtual reality environment. The subject sees virtual objects rendered in the head-mounted display, and grasps these objects using the force-feedback devices connected to his thumb and index fingers.

visual information was highly "noisy"? The experimental data collected by Jacobs and his colleagues indicates that people start averaging the locations indicated by visual and auditory modalities, with the auditory location being assigned a larger weight when the visual information is corrupted with noise. These data can be accounted for by an optimal statistical model known as a Kalman filter, indicating that people in these studies integrated the information provided by their visual and auditory modalities in an optimal manner.

More recently, Jacobs has been examining whether human perceptual learning is optimal. Imagine that each time you saw and grasped an object, your (unconscious) estimate of the object's depth based on haptic (touch) information was always 10% larger than your estimate based on visual information. If you had reason to believe that your haptic estimates were more reliable, would you adapt your visual system so that, in the future, its interpretations were more consistent with those of your haptic system? Jacobs and his colleagues used a novel virtual reality environment to address this question (see photo on opposite page). People in their experiments wore a head-mounted display which allowed them to see "virtual" objects. In addition, people's thumbs and index fingers were connected to robot arms comprising a haptic force-feedback device which allowed them to grasp these objects. The experimental data indicate that when there is a discrepancy between people's visual and haptic percepts, and when the visual information is noisy, people recalibrate their visual systems so that these systems' interpretations are more consistent with those of their haptic systems. These data support the hypothesis that people continuously monitor the interpretations of their sensory systems, and keep them calibrated by adapting each system whenever its interpretations fail to match the interpretations of other systems.

Jacobs continues to examine the optimality of people's perceptual learning. "When quantitatively comparing people's learning performances with those of state-of-the-art machine learning devices using the same data sets, it often seems as if people do not learn as much from each data item as they theoretically could," adds Jacobs. "Why do people show sub-optimal learning?"

Jacobs is now examining different training procedures to see which ones elicit the largest learning effects. Finding effective training procedures for perceptual skills has enormous importance for science education. For example, radiologists need to learn to visually distinguish tumors from other tissue by viewing mammograms, geologists need to learn to visually recognize different types of rock samples, astronomers need to learn to visually interpret stellar spectrograms, and biologists need to learn to visually identify different cell structures. Although much is known about how to train people to learn new facts, defining effective training procedures for teaching people new perceptual skills is an open area of research.

Ph.D. alums **Peter Desnoyers** and **Charles Sutton** received the 2007-2008 Awards for an Outstanding Doctoral Dissertation, sponsored by Yahoo!. Desnoyer's dissertation was titled "Distributed Sensing Networks: Archiving, Indexing, and Querying," and Sutton's was "Local Training for Conditional Random Fields." Both received their Ph.D.s in 2008.

IEEE Intelligent Systems selected **Jennifer Neville** (Ph. D. '06) as one of "AI's 10 to Watch." The feature on the 10 top researchers was in the May/June issue. Neville is an Assistant Professor of Computer Science and Statistics at Purdue University. Her research interests are in machine learning and data mining.

Amy McGovern (Ph.D. '02) received a five year NSF CAREER grant for her project "Developing Dynamic Relational Models to Anticipate Tornado Formation." McGovern is an Assistant Professor in the School of Computer Science at the University of Oklahoma.

Balaraman ("Ravi") Ravindran (Ph.D. '04) was awarded tenure and a promotion to Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Madras.

Center for Intelligent Information Retrieval alums **Jeremy Pickens** (Ph.D. '04) and **Chirag Shah** (M.S. '06) along with co-authors Gene Golovchinsky, Pernilla Qvarfordt, and Maribeth Back received the Best Paper Award at the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08) for their paper "Algorithmic Mediation for Collaborative Exploratory Search." Pickens is a Research Scientist at FX Palo Alto Labs and Shah is a Ph.D. student at the University of North Carolina, Chapel Hill.

Outstanding CS alum awards

Department Chair Andy Barto has announced a new Awards Program aimed at recognizing outstanding achievements by alums of the department. Awards will be presented at an annual UMass Amherst Computer Science Awards Banquet; the first is planned for the weekend of May 1, 2009. The awards will recognize achievement in such areas as entrepreneurship, scientific research, and education.

Details such as award categories, the nomination process, and the date, place, and time of the awards banquet, will be forthcoming. CS Alums are encouraged to stay in touch and participate by making a nomination and/or attending the awards banquet.

Barto also announced steps to make it easier for CS Alums to keep up with department news. In addition to the biannual Significant Bits newsletter, and the web site, www.cs.umass.edu, the department has also set up a LinkedIn group (UMass Amherst Dept. of Computer Science), and a Facebook group (UMass CS) that all CS Alums are encouraged to join. The LinkedIn and Facebook groups are expected to be a forum for discussions and news of particular interest to CS Alums. Be sure to check them out!

Ron Bekkerman: *Combinatorial Markov Random Fields and their Applications to Information Organization* (James Allan, Advisor); Feb. 2008; Research Scientist, HP Labs.

We propose a new type of undirected graphical model called a Combinatorial Markov Random Field (Comraf). An efficient inference methodology for Comrafs based on combinatorial optimization of information-theoretic objective functions; both global and local optimization schema are discussed. We apply Comrafs to multi-modal clustering tasks: standard (unsupervised) clustering, semi-supervised clustering, interactive clustering, and one-class clustering. For the one-class clustering task, we analytically show that the proposed optimization method is optimal under certain simplifying assumptions. We empirically demonstrate the power of Comraf models by comparing them to other state-of-the-art machine learning techniques, both in text clustering and image clustering domains. For unsupervised clustering, we show that Comrafs consistently and significantly outperform three previous state-of-the-art clustering techniques on six real-world textual datasets. For semi-supervised clustering, we show that the Comraf model is superior to a well-known constrained optimization method. For interactive clustering, Comraf obtains higher accuracy than a Support Vector Machine, trained on a large amount of labeled data. For one-class clustering, Comrafs demonstrate superior performance over two previously proposed methods. A comprehensive recipe for machine learning modeling with Comrafs is given.

Aron Culotta: *Learning and Inference in Weighted Logic with Application to Natural Language Processing* (Andrew McCallum, Advisor); May 2008; Assistant Professor, Southeastern Louisiana University.

Statistical machine learning approaches to natural language processing have largely replaced earlier logic-based systems. These probabilistic methods are well-suited to the ambiguity inherent in human communication. The shift to statistical modeling has mostly abandoned the representational advantages of logic-based approaches. Although many language processing problems can be more meaningfully expressed in first-order logic. Unfortunately, most machine learning algorithms have been developed for propositional knowledge representations.

In recent years, there have been a number of attempts to combine logical and probabilistic approaches to artificial intelligence. However, their impact on real-world applications has been limited because of serious scalability issues that arise when algorithms designed for propositional representations are applied to First-order logic representations. In this thesis, we explore approximate learning and inference algorithms that are tailored for higher-order representations, and demonstrate that this synthesis of probability and logic can signifcantly improve the accuracy of several language processing systems.

Peter Desnoyers: *Distributed Sensing Networks: Archiving, Indexing, and Querying* (Prashant Shenoy, Advisor); Feb. 2008; Assistant Professor, Northeastern University.

This thesis advocates data-aware and storage-centric approaches to data handling in distributed data collection environments. Raw data is stored locally, and only selected, relevant information reaches the application. We apply two strategies: (i) archiving, indexing and querying to retrieve discrete portions of data, and (ii) mathematical modeling to extract relationships spread diffusely across the data. We propose mechanisms for these strategies and evaluate them in appropriate system contexts.

First we address storage and indexing of high-speed event data. A disk-based storage system guarantees high write rates, and a signature file-based index to index high-speed data in real time. We evaluate a network monitor using these mechanisms.

Next, we examine indexing and retrieval in low-power wireless sensor networks. Sensors send data summaries to more powerful proxy nodes, which use a novel index structure, the Interval Skip Graph, representing multiple records by a single imprecise key. We present a prototype sensor network storage system using these mechanisms.

Lastly we address analysis of feature-rich but poorly structured events. Statistical machine learning techniques create models of application behavior in a data center, using automated feature identification to identify model inputs within the raw data. We present and evaluate Modellus, a monitoring system using these mechanisms.

Fernando Diaz: *Regularizing Query-Based Retrieval Scores* (James Allan, Advisor); Feb. 2008; Research Scientist, Yahoo!, Montreal.

Query-based information retrieval refers to the process of scoring documents given a short natural language query. Query-based information retrieval systems have been developed to support searching diverse collections such as the world wide web, personal email archives, news corpora, and legal collections. This thesis is motivated by one of the tenets of information retrieval: the cluster hypothesis. We define a design principle based on the cluster hypothesis which states that retrieval scores should be locally consistent. We refer to this design principle as score autocorrelation. Our experiments show that the degree to which retrieval scores satisfy this design principle correlates positively with system performance. We define a general, black box method, local score regularization, for improving the local consistency of a set of retrieval scores. We demonstrate that regularization consistently and significantly improves retrieval performance for a wide variety of baseline algorithms. Regularization is closely related to classic techniques such as pseudo-relevance feedback and cluster-based retrieval. We demonstrate that the effectiveness of these techniques may be explained by their regularizing behavior. We argue that regularization should be adopted either as a generic post-processing step or as a fundamental design principle for retrieval models.

Shaolei Feng: *Statistical Models for Text Query-Based Image Retrieval* (R. Manmatha, Advisor); May 2008; Research Scientist, Siemens Corporate Research.

Image indexing and retrieval is a challenging problem. Traditional content-based approaches make use of queries based on image examples or image attributes like color and texture. However, they do not capture the semantics or meanings of images. Generally, libraries and other organizations manually annotate each image with keywords and captions. The disadvantage of this approach is the huge cost of annotating a large number of images and the inconsistency of annotations by different people. In this work, we focus on general image and historical handwritten document retrieval based on textual queries. We explore three statistical model-based techniques that allow us to retrieve general images and historical handwritten document images with text queries: (i) image retrieval based on automatic annotation, (ii) direct

retrieval based on computing the posterior of an image given a text query, and (iii) handwritten document image recognition.

The contributions of this work include (i) two probabilistic generative models for annotation-based retrieval, (ii) a direct retrieval model for general images, and (iii) an investigation of machine learning models for handwritten document recognition. Our experimental results show that our approaches may be applied to practical textual query-based retrieval systems on large image data sets.

Jiwoon Jeon: *Searching Question and Answer Archives* (W. Bruce Croft, Advisor); Sept. 2007; Software Engineer, Google Inc.

Archives of questions and answers are a valuable information source. However, little research has been done to exploit them. We propose a new type of information retrieval system that answers users' questions by searching question and answer archives. The proposed system has many advantages over current web search engines. In this system, natural language questions are used instead of keyword queries, and the system directly returns answers instead of lists of documents. Two most important challenges in the implementation of the system are finding semantically similar questions to the user question and estimating the quality of answers. We propose using a translation-based retrieval model to overcome the word mismatch problem between questions. Our model combines the advantages of the IBM machine translation model and the query likelihood language model and shows significantly improved retrieval performance over the state of the art retrieval models. We also show that collections of question and answer pairs are good linguistic resources for learning reliable word-to-word translation relationships. To avoid returning bad answers to users, we build an answer quality predictor based on statistical machine learning techniques. By combining the quality predictor with the translation-based retrieval model, our system successfully returns relevant and high quality answers to the user.

Giridhar Kumaran: *Efficient User Interaction In Information Retrieval* (James Allan, Advisor); May 2008; Scientist, Microsoft Live Labs.

We present new ways of interacting with a user based on query analysis and reformulation. We do this by showing users the potential impact their decisions will have on the retrieval process, so they can make more informed choices from the options presented to them.

Unlike a one-procedure-fits-all strategy, we invoke user interaction only when there is potential for improvement because tedious interaction can have an unfavorable impact on user experience. We present techniques for selective interaction and show their utility in the context of two interaction techniques we have developed. Our results show that user interaction can be avoided in most cases without much deterioration in performance.

User interaction can be made more productive by providing users with an optimally-sized set of high quality options. We present efficient techniques to determine such a set. When faced with a decision to interact with a user given a query, we determine the best interaction technique for that query by obtaining implicit feedback from the user. By utilizing these interaction-related techniques, we show through simulations and user studies that users can obtain better performance with less effort.

Audrey Lee: *Geometric Constraint Systems with Applications in CAD and Biology* (Ileana Streinu, Advisor); May 2008; Visiting Prof., Dept. of CS, Mount Holyoke College.

Motivated by applications in Computer Aided Design (CAD) and biology, we investigate geometric constraint systems, composed of atomic elements with constraints between them. A well-studied model in rigidity theory is the bar-and-joint structure, where the atomic elements are universal joints connected by fixed-length bar constraints. We propose several new models involving constraints arising in CAD and biology and provide the theoretical foundation for each. In particular, we present a model addressing the pairwise constraints among points, lines and planes found in constraint-based CAD software, such as in the assembly environment of the widely-used SolidWorks CAD application.

As a result, we identify and generalize combinatorial properties that appear as necessary conditions for generic rigidity of these new constraint systems; in some cases, the conditions are also sufficient, thus providing a complete characterization. We study sparsity for graphs arising from known rigidity results and present extensions of this concept to graded, mixed and nested sparsity. For these sparsity properties, we present algorithms that solve the fundamental questions of Decision, Extraction and Components, based on the efficient and elegant pebble games first developed for planar bar-and-joint rigidity.

Wei Li: *Pachinko Allocation: DAG-Structured Mixture Models of Topic Correlations* (Andrew McCallum, Advisor); Sept. 2007; Senior Software Development Engineer, Yahoo! Inc.

Statistical topic models are increasingly popular tools for summarization and manifold discovery in discrete data. However, the majority of existing approaches capture no or limited correlations between topics. We propose the pachinko allocation model (PAM), which captures arbitrary, nested, and possibly sparse correlations between topics using a directed acyclic graph (DAG). We present various structures within this framework, different parameterizations of topic distributions, and an extension to capture dynamic patterns of topic correlations. We also introduce a nonparametric Bayesian prior to automatically learn the topic structure from data. The model is evaluated on document classification, likelihood of held-out data, the ability to support fine-grained topics, and topical keyword coherence. With a highly-scalable approximation, PAM has also been applied to discover topic hierarchies in very large datasets.

Marc Liberatore: *Low-Latency Anonymity Systems: Statistical Attacks and New Applications* (Brian Levine, Advisor); Feb. 2008; Postdoctoral Fellow/Visiting Faculty, Dept. of Math and CS, Wesleyan University.

In this dissertation, we study low-latency anonymity protocols and systems. Such systems enable anonymous communication where latency is not tolerated well, such as browsing the web, but also introduce new vulnerabilities not present in systems that hide timing information. We examine one such vulnerability, the profiling attack, as well as possible defenses. We also examine the feasibility of using low-latency anonymity techniques to support a new application, Voice over IP (VoIP). First, we show that profiling attacks on low-latency anonymity systems are feasible. The attack is based upon preconstructing profiles of communication, and using them to identify the sender of encrypted, anonymized traffic. Second,

we present results from a large-scale measurement study, which indicate that profiling is practical across sets of thousands of possible senders, and that such profiles remain valid for weeks at a time. Third, we evaluate defenses against the profiling attack and their effects upon system performance. We then demonstrate the feasibility of supporting anonymous VoIP; specifically, we show supporting measurement data and outline the changes current anonymity systems would require. We also show how such systems are potentially more vulnerable to known attacks, and examine the tradeoffs between VoIP performance and anonymity inherent in such systems.

Junning Liu: *On Joint Coding and Combinatorial Optimization in Communication Networks* (Donald Towsley and Micah Adler, Advisors); Feb. 2008; R & D Software Engineer, YouTube Analytics, a subsidiary of Google.

Information flows in communication networks are treated as commodity flows. Network coding is a technique that allows nodes to manipulate the information content arbitrarily inside the network. This work focuses on the joint coding/routing optimization for throughput, energy consumption, and robustness of wired and wireless networks. We strive to minimize communication costs of collecting correlated data through a network at a sink, how to maximize the throughput of multi-pair independent unicast traffic, and how to maximize the data utility in random, unreliable data collection networks. We introduce "distance entropy," which characterizes the spatial distribution of information at networked sources; we show that it is a meaningful metric as the tight lower bound of the minimum communication cost for collecting the sources at a single sink. We propose a practical coding/ routing algorithm called Hierarchical Difference Broadcasting (HDB). We also revisit Gupta & Kumar's work on the capacity of wireless ad hoc networks and consider it in a setting that allows arbitrary coding at the nodes. We characterize the throughput capacity order and show coding combined with wireless broadcasting provides no order difference improvement. Finally, we apply coding to improve the network robustness and analyze a joint coding & scheduling problem.

Xiaoyong Liu: *Cluster-Based Retrieval from the Language Modeling Perspective* (W. Bruce Croft, Advisor); Feb. 2008; Software Design Engineer, Hewlett Packard.

It is generally assumed that the relevance of documents can be assessed independently. The fact that a document is relevant does not contribute to predicting the relevance of a closely-related document. In contrast, cluster-based retrieval assumes that the probability of relevance of a document should depend on the relevance of other similar documents to the same query.

The most common approach to cluster-based retrieval is to retrieve one or more clusters in their entirety. Research in this area suggests that optimal clusters could yield very large improvements in effectiveness relative to document retrieval. However, no retrieval strategy has achieved this result; document retrieval is generally more effective. One area of recent research is to use clusters as a form of document smoothing.

We retrieve the best group of documents, from the language-modeling perspective. We study both cluster smoothing and cluster retrieval. We analyze the advantages and disadvantages of a range of representation techniques, derive features that characterize good document clusters, and develop new probabilistic representations that capture the identified features. An extensive empirical evaluation is provided. We argue that the use of good document clusters by an IR system depends on how they are represented.

Donald Metzler: *Effectively Modeling Term Dependencies in Information Retrieval* (W. Bruce Croft, Advisor); Sept. 2007; Research Scientist, Yahoo! Research.

Current state of the art information retrieval models treat documents and queries as bags of words. There have been many attempts to go beyond this simple representation. Unfortunately, few have shown consistent improvements in retrieval effectiveness across a wide range of tasks and data sets. Here, we propose a new statistical model for information retrieval based on Markov random fields. The proposed model goes beyond the bag of words assumption by allowing dependencies between terms to be incorporated into the model. This allows for a variety of textual and non-textual features to be easily combined under the umbrella of a single model. Within this framework, we explore the theoretical issues involved, parameter estimation, feature selection, and query expansion. We give experimental results from a number of information retrieval tasks, such as ad hoc retrieval and web search.

Trevor Strohman: *Efficient Processing of Complex Features for Information Retrieval* (W. Bruce Croft, Advisor); Feb. 2008; Software Engineer, Google Inc.

Text search systems research has primarily focused on simple occurrences of query terms within documents to compute document relevance scores. However, recent research shows that additional document features are crucial for improving retrieval effectiveness. We develop a series of techniques for efficiently processing queries with feature-based models. Our TupleFlow framework, an extension of MapReduce, provides a basis for custom binned indexes, which efficiently store feature data. Our work in binning probabilities shows how to effectively map language model probabilities into the space of small positive integers, which helps improve speeds without reducing query effectiveness. We also show new efficient query processing results for both document-sorted and score-sorted indexes. All of our work is evaluated using the largest available research dataset.

Kyoungwon Suh: *Monitoring, Measurement, and Control of Multimedia Traffic in IP Networks* (James F. Kurose and Donald F. Towsley, Advisors); Sept. 2007; Assistant Professor, Illinois State University at Normal.

We propose several architectural components that monitor and measure multimedia traffic at the edge and in the core of IP networks and serve to clients in edge networks. We present a technique to detect Skype-relayed traffic from passively measured packet traces collected at the edge of the Internet. We propose several flow-level metrics to characterize the nature of relayed traffic which is then used to detect Skype-relayed traffic. We consider the problem of optimal placement within the network and sampling rates for the monitors. We formulate several minimum-cost, maximum-coverage problems under various budget constraints and show that they are NP-hard. We propose and evaluate greedy heuristics. Last, we introduce and evaluate the Push-to-Peer architecture for streaming video among cooperating nodes in an edge network, that can drastically reduce the load posed on the core and access links of the network and on the streaming servers. The main departure from previous designs is that content is proactively pushed to peers and persistently stored before the actual peer-to-peer transfers.

Charles Sutton: *Local Training for Conditional Random Fields* (Andrew McCallum, Advisor); Feb. 2008; Postdoctoral Researcher, Department of Electrical Engineering and Computer Science, University of California Berkeley.

Many applications require predicting multiple interdependent variables. Recent attention has therefore focused on structured prediction methods, which combine the modeling flexibility of graphical models with the complex, dependent features of traditional classification methods. Especially popular have been conditional random fields (CRFs)—graphical models of the conditional distribution over outputs given a set of observed features. Unfortunately, parameter estimation in CRFs requires repeated inference, which can be computationally expensive.

I investigate efficient training methods for CRFs with complex graphical structure, focusing on local methods to avoid propagating information globally. First, I investigate piecewise training, which trains each of a model's factors separately. I present three views of piecewise training: as maximizing the likelihood in a so-called "node-split graph", as maximizing the Bethe likelihood with uniform messages, and as generalizing the pseudo-moment matching estimator of Wainwright et al (2003). Second, I propose piecewise pseudolikelihood, a hybrid procedure which "pseudolikelihood-izes" the piecewise likelihood, and is more efficient if the variables have large cardinality. Piecewise pseudolikelihood performs well even on applications in which standard pseudolikelihood performs poorly. Finally, motivated by the connection between piecewise training and BP, I explore training methods using beliefs arising from stopping BP before convergence.

Xing Wei: *Topic Models in Information Retrieval* (W. Bruce Croft, Advisor); Sept. 2007; Research Scientist, Yahoo! Inc.

Topic modeling demonstrates the semantic relations among words, which should be helpful for information retrieval tasks. We present probability mixture modeling and term modeling methods to integrate topic models into language modeling framework for information retrieval. A variety of topic modeling techniques, including manually-built query models, term similarity measures and latent mixture models, especially Latent Dirichlet Allocation (LDA), a formal generative latent mixture model of documents, have been proposed or introduced into IR tasks. We investigated and evaluated them on several TREC collections within presented frameworks, and show that significant improvements over previous work can be obtained. Practical problems such as efficiency and scaling considerations are discussed and compared for different topic models. Other recent topic modeling techniques are also discussed.

Jerod Weinman: *Unified Detection and Recognition for Reading Text in Scene Images* (Allen Hanson and Erik Learned-Miller, Advisors); May 2008; Assistant Professor, Department of Computer Science, Grinnell College.

Computers can currently "read" document text for the blind about as well as a seven-year-old. Scene text recognition brings new challenges. A central limitation of current approaches is a feed-forward, bottom-up, pipelined architecture that isolates the many tasks and information involved in reading. The result is a system that commits irreversible errors and has components that lack access to relevant information.

We propose a system for scene text reading that in its design, training, and operation is more integrated, which could lead to improved performance. First, we present a simple contextual model for text detection that is ignorant of any recognition. Through the use of special features and data context, this model performs well on the detection task, but limitations remain due to the lack of interpretation. We then introduce a recognition model that integrates several information sources, including font consistency and a lexicon, and compare it to pipelined approaches. Next we examine a unified framework where features are selected for the joint detection and recognition task, yielding better results with fewer features. Finally, we demonstrate a model that integrates segmentation and recognition of both words and characters to recognize text with difficult layouts and low resolution more accurately.

Xiaolan Zhang: *Routing in DTN: Performance Modeling, Network Coding Benefit and Real Trace Studies* (James Kurose and Donald Towsley, Advisors); Sept. 2007; Assistant Professor, Fordham University.

We study three problems on unicast routing in Disruption-Tolerant Networks (DTNs). First, we propose a unified framework based on Ordinary Differential Equations (ODEs) to study the performance of a class of epidemic style routing schemes. Derived as the limit of Markov process model under proper scaling, the ODE models capture the propagation and recovery process of data packets. We derive a rich set of closed-form results using the ODE models, and quantitatively characterize their performance.

Next, we investigate the benefit of applying Random Linear Coding (RLC) in resource-constrained DTNs. We explore different ways to apply network coding and study cases with different network traffic load. We show that the RLC-based scheme achieves minimal block delivery delay with high probability and improves the tradeoff between block delivery delay and number of transmissions.

Last, we study mobility traces taken from UMass DieselNet. We analyze the bus-to-bus contact traces to develop a generative model and then generate synthetic traces. We show that an aggregate model for the inter-contact time is too coarse to accurately capture DTN routing performance. We construct a route-level inter-contact time model, which captures interesting mobility structure within the trace.

Yun Zhou: *Retrieval Performance Prediction and Document Quality* (W. Bruce Croft, Advisor); Feb. 2008; Software Engineer, Google.

The ability to predict retrieval performance has potential applications in many important IR (Information Retrieval) areas. In this thesis, we study the problem of predicting retrieval quality at the granularity of both the retrieved document set as a whole and individual retrieved documents. At the level of ranked lists of documents, we propose several novel prediction models that capture different aspects of the retrieval process that have a major impact on retrieval effectiveness. These techniques make performance prediction both effective and efficient in various retrieval settings including a Web search environment. As an application, we also provide a framework to address the problem of query expansion prediction. At the level of documents, we predict the quality of documents in the context of Web ad-hoc retrieval. We explore document features that are predictive of quality. Furthermore, we propose a document quality language model to improve retrieval effectiveness by incorporating quality information.

Significant Bits

Newsletter of the
Department of Computer Science
College of Natural Sciences & Mathematics
at the University of Massachusetts Amherst

140 Governors Drive
University of Massachusetts Amherst
Amherst, MA 01003-9264

"Significant Bits" is published twice a year by the Department of Computer Science, University of Massachusetts Amherst **(www.cs.umass.edu)**.
Your news, suggestions, comments, and contributions are welcome. Please mail them to the address above or send them electronically to **bits@cs.umass.edu**.

Department Chair	Andrew Barto
Editor	Jean Joyce
Art Direction	North Haven Design
Faculty Liaisons	Bruce Croft, Mark Corner, Gerome Miklau
Graduate Student Liaisons	Filip Jagodzinski, Borislava Simidchieva
Contributors	Rick Adrion, Michelle Eddy, Robert Jacobs, Laura Macsuga, Paul Sihvonen-Binder (photos), UMass Amherst News Office staff

NONPROFIT ORG.
U.S. Postage
PAID
Permit No. 2
AMHERST, MA

Thanks for your support

The following alumni and friends have actively supported the Department of Computer Science from April 2008 through September 2008. Such financial support is greatly appreciated and helps maintain a world-class instructional and research program. Contributions from alums and friends help to fund important special activities that are not supported through the state budget.

Those interested in helping the department should visit **www.cs.umass.edu/about/donate** for online donations or send a check made out to UMass Amherst to: Department of Computer Science, Attn: Jean Joyce, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA 01003-9264

Please state that your gift is restricted to Computer Science.

Mr. John F. Adler (1985)
Mr. Vishal Dhiman & Ms. Surbhi Agarwal (2004)
Mr. Donald Bray
Mrs. Mary C. Brown (1985)
Dr. Brendon Cahoon (2002)
Ms. Lesley Cederlund (1985)
Jennifer & Thomas Chistolini
Mr. Eric Christophersen (2005)
Janet & Luis Chukiu
Mr. Stephen M. Werntz & Mrs. Kimberly Edgar (1986)
Mr. Craig L. Fournier (1991)
Mr. Kenneth E. Groder, III (1980)
Dr. Oscar J. Gonzalez (2001)
Mr. John Goshdigian (1973)
Mr. Alexander J. Harm (1996)
Mr. Michael J. Hartman (1982)
Mr. Robert L. Hazen (1983)
Mr. Rajesh K. Jha (1990)
Dr. Wei Li (2007)
Mr. Gregory J. Lorge (2006)
Karen & Andrew Merlino, Jr. (1986)
Mr. Michael C. Monks (1987)
Dr. Jitendra Padhye
Dr. Achilleas Papakostas (1993)
Ms. Laura Anne Price (1987)
Mr. Thomas M. Reilly (1986)
Mr. Gilbert D. Ross Jr. (1982)
Dr. Christina Fong & Dr. Tuomas Sandholm (1986)
Mr. Michael J. Scudder (1990)
Mr. Daniel W. Slowik (1971)
Kathleen & Walter Stover, Jr.
Mr. Michael J. Sullivan (1971)
Amy & Keith Swan (1990)
Mr. Robert P. Thomas (1984)
Mr. Matthew W. Thompson (2007)
Mr. Matthew Van Laethem (2003)
Tsong-Shyan & Min-Jen Wang (1987)
Dr. Zhenlin Wang (1999)
Mrs. Nancy J. Weinreich (1983)
Mr. John R. Welch (1994)
Dr. Charles J. Welty (1979)
Mr. Peter F. Willey (1980)

Corporate and Matching Gifts
Cisco Systems
EMC Corporation
Google
IBM Intl. Foundation
McKesson
Microsoft
Nokia Research Center
Springfield Neurology
Sun Microsystems
Yahoo!

We need your continued support

Gifts like yours help the department in many ways, such as funding departmental seminars by outstanding scientists, assisting undergraduate research, and helping new faculty establish their research programs. In addition to contributions ear-marked for a specific purpose, general support helps make it possible for us to continue activities that enrich our educational and research programs.

A current need within the department is to upgrade the equipment in our instructional education lab and PC lab with new computers. The department also needs to furnish the newly renovated study and classroom space in the Lederle Lowrise. In addition, donations to our CS Endowment fund will have continuing benefits to the department's graduate and undergraduate programs.