University of Massachusetts Amherst

Fall / Winter 2009

Significant BITS

Newsletter of the
Department of Computer Science

Fu named *Technology Review's* 2009 Innovator of the Year

Technology Review recognized Assistant Professor Kevin Fu as one of the magazine's 2009 Young Innovators Under Age 35 (TR35), an elite group of accomplished young innovators who exemplify the spirit of innovation. Chosen from among this year's TR35 winners, Fu was honored with the 2009 Innovator of the Year Award for his cutting-edge research to improve computer security and privacy for such applications as implantable medical devices and contactless

continued on page 3

News

Lesser named Distinguished Professor 4
High Performance Computing Center Planned 7
Adjunct Professors named 11

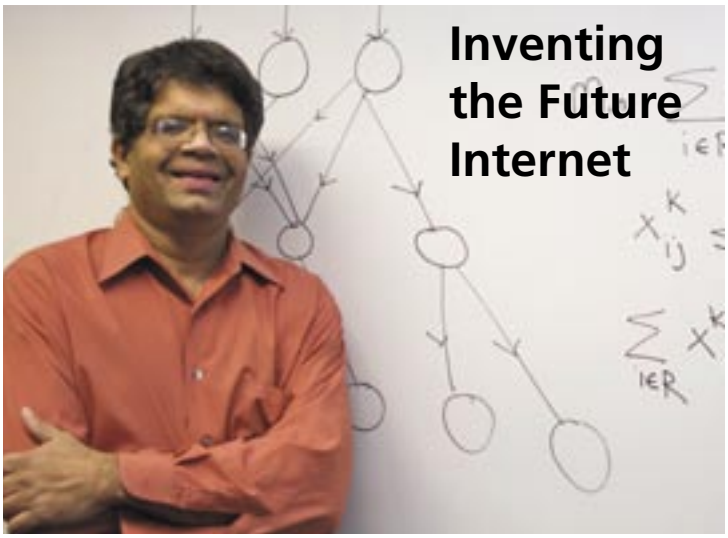
Awards

CNS Honors Clarke 5
Moss receives Chancellor's Award 14

Alums

Alum Focus: Bryant York 8
Recent Ph.D. Graduates 10

Inventing the Future Internet



Ramesh Sitaraman

Roy Plunkett discovered Teflon seventy years ago when he was researching new refrigerants, a far cry from the non-stick pans, computer chips, and space suits where Teflon is used today. Serendipity plays a greater role in technological research than people suspect! Perhaps it is no surprise that the Internet is now largely used in ways not envisioned by its original architects. The Internet is transforming every aspect of human society by enabling a wide range of applications for business, commerce, entertainment, news, and social networking. Can the Internet realize its potential and effectively host the novel applications that are transforming our future? "Not without new scientific breakthroughs," says Associate Professor Ramesh Sitaraman, whose research is focused on this key question from a foundational perspective.

Modern and future distributed applications require high reliability, performance, security, and scalability, but require low operating costs. For instance, major e-commerce sites require at least 99.99% reliability, allowing no more than a few minutes of downtime per month! As another example, the future migration of high-quality television to the Internet would require the ability to flawlessly transport tens of petabits of data per second to audiences around the world. However, the Internet was never designed to provide the stringent requirements of such modern and futuristic distributed applications. "The Internet is a vast patchwork of more than 13,000 autonomous networks that often compete," says Sitaraman. "Transporting information across this patchwork can result in failures and performance degradation, especially at the transit and peering points where the networks exchange traffic. A clean-slate redesign of the Internet is appealing, but would be hard to implement given the wide-spread adoption of the current technology." So, what options do we have left? The age-old concept of virtualization suggests an approach. We can build a *virtual* network called a Content Delivery Network (CDN) over the existing Internet to provide an *abstraction* of the stringent requirements that applications need. In essence, a CDN is an extremely large distributed system of servers that run sophisticated distributed algorithms to coordinate the delivery of web content, streaming media, and online applications to end-users, with much greater reliability, performance, security, and scalability than the underlying "vanilla" Internet.

Sitaraman's interest in virtual networks started with his doctoral work at Princeton in 1993 when he explored techniques for creating virtual networks that provide the abstraction of being fully-reliable and fault-free, even though they were built over

continued on page 4


Andrew Barto

New College, New Connections

At the start of the Fall 2009 semester, the Department of Computer Science found itself in a new college. From one of nine departments in the former College of Natural Sciences and Mathematics (NSM), we became one of fifteen departments in the new College of Natural Sciences (CNS). This new college was formed by combining the departments of NSM with most of the departments from the College of Natural Resources and the Environment (NRE) together with the Department of Psychology, formerly in the College of Social and Behavioral Sciences. This new organization brings together departments with common interests under the roof of a single college. In particular, it brings together researchers in the life, environmental, computational, and physical sciences. Being part of the new CNS will encourage new connections between faculty and students in a wide range of disciplines and will enhance our ability to initiate and manage new multi-disciplinary projects.

Connections across sub-disciplines *within* Computer Science has long been a hallmark of the department. We are proud of this tradition and work hard to maintain it. Looking at our *graph of research interactions* on our web site, you will see a densely connected web of research relationships within the department. However, another hallmark of the department is the web of collaborative relationships that connect us with scholars in many disciplines across the campus and at other institutions. The within-department interaction graph shows only a small fraction of the collaborations in which our faculty members routinely engage.

The breadth and strength of our connections across the campus were made very evident last Spring by a challenge from the UMass campus leadership. Departments were summoned to propose clusters of new faculty hires that could form the nuclei of groundbreaking interdisciplinary research and teaching efforts. Given the dire impact of the economic situation on the campus, this process was advanced as a way to maximize the impact of new faculty hires, however few that number may turn out to be. The Computer Science faculty responded to this challenge with impressive energy, in part because many members of our faculty were already engaged in projects that reach beyond the department, and in part because Computer Science is such an integral part of progress in so many other fields.

The department participated in more proposals for clusters of new faculty positions than any other department on campus. The proposals included the following: **Systems and Computational Biology,** with Biochemistry and Molecular Biology, Microbiology, Biology, and Mathematics and Statistics; **High Performance Green Computing,** with Astronomy; **ASSISTHealth,** with Nursing and Mechanical and Industrial Engineering; **Language, Experimentation and Computation,** with Linguistics and Psychology; **Judgment and Decision-Making under Uncertainty,** with Psychology and Economics; **Computational Social Science,** with Political Science, Sociology, and Mathematics and Statistics; and **Cyber-security,** with Electrical and Computer Engineering, Finance and Operations Management, Communication, and Political Science. Almost every unit on campus was included in this group of proposals.

As I write this, I do not know which, if any, of these proposals will yield new faculty positions, but the range of multi-disciplinary directions represented by these proposals—many built on existing thriving collaborations—provide ample evidence that our graph of research interactions includes a great many edges going outward from the department to connect widely across other disciplines. As we move forward, we will be putting great energy into deepening the department in traditional core areas of Computer Science. But we will also be widening the department by adding more faculty members with multi-disciplinary expertise to help us meet new challenges and explore new frontiers of inquiry. Computer Science and other disciplines are converging in unexpected ways that promise exciting advances over the future. UMass Amherst Computer Science will be a very active participant in these advances, and in doing so will be a leader in what is to come.

Shenoy named ACM Distinguished Member

Professor Prashant Shenoy was chosen as a Distinguished Member of the Association for Computing Machinery (ACM). The Distinguished Member grade recognizes those ACM members with at least 15 years of professional experience and 5 years of continuous Professional Membership who have achieved significant accomplishments or have made a significant impact on the computing field.

Shenoy's research interests lie in operating and distributed systems, sensor networks, Internet systems, and multimedia. His current research projects focus on server and desktop virtualization, RFID networks, sensor data management, and pervasive multimedia systems.

He received a B.Tech in Computer Science and Engineering from the Indian Institute of Technology, Bombay, in 1993 and his M.S. and Ph.D. in Computer Science from the University of Texas at Austin in 1994 and 1998, respectively. Shenoy joined the department's faculty in 1998 and was promoted in 2009 to full Professor. He heads the Laboratory for Advanced Systems Software.

Fu receives NSF CAREER Award

Assistant Professor Kevin Fu received a five-year National Science Foundation (NSF) Faculty Early Career Development (CAREER) award for his project "Computational RFID (radio-frequency identification) for Securing Zero-Power Pervasive Devices."

Fu's research project advances the knowledge and capability for securely performing general-purpose computation and communication on zero-power devices that rely on harvested energy stored in extremely small reservoirs such as capacitors. "A type of zero-power device, computational RFIDs (CRFIDs) harvest RF energy and endure continual interruptions to power," says Fu. Complete loss of RAM on the order of every second makes the notion of a computational checkpoint fundamental to this model of computing. A significant problem is how to perform effective computation and checkpoints in a secure and privacy-preserving manner. Fu adds, the most fundamental question is: how to securely make forward progress of computation on zero-power devices? The exploration of checkpointing strategies that exploit the energy-centric properties of computational RFIDs will lay a solid foundation for designing secure software and systems for pervasive devices.

Beyond security and privacy, the research seeks to discover what kinds of computational problems can be solved in a practical sense on extremely resource-constrained devices. While the experiments focus on computational RFIDs, many of the techniques will apply broadly to devices with batteries as well. This work will lead to improved security for computation in low-power, untrusted infrastructure. Potential applications include improved security and privacy for implantable medical devices and sensors embedded in concrete components of bridges and roads.

After receiving his Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Fu joined the UMass Amherst Computer Science faculty as an Assistant Professor in 2005. He leads the Security and Privacy Research group (SPQR) at UMass Amherst. He is the director of the RFID Consortium on Security and Privacy and the co-director of the Medical Device Security Center.

The CAREER program offers the NSF's most prestigious awards for new faculty members. It 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. In addition to Fu, sixteen members of the department's faculty have received previous CAREER awards.

FU INNOVATOR OF THE YEAR

– – *continued from page 1*

credit cards without compromising safety or effectiveness.

Fu, a member of the CS Department faculty since 2005, was selected from more than 300 nominees by a panel of expert judges and the *Technology Review* editorial staff. Fu is the first researcher from the UMass system to be so honored by the technology magazine, which calls itself the authority on the future of technology.

In 2008, Fu's research team showed that implantable heart defibrillators and similar devices are vulnerable to hacking, which led Fu to begin designing and testing, both in laboratory and clinical settings, zero-power technology and low-power cryptographic protocols for implantable medical devices such as heart defibrillators and pacemakers. Zero-power means the tiny chips will run without draining the device's batteries, making devices more effective and secure.

Modern implantable medical devices rely on radio communication for diagnostic and therapeutic functions, allowing health care providers to remotely care for patients. However, using radio communication and the Internet in a medical setting comes with more significant security and privacy risk than does desktop computing, for example. Further, it's more difficult to protect the privacy of sensitive information on implantable medical devices than it is for electronic health records or pharmacy databases, Fu says.

This work comes at a critical time because emerging medical devices will share sensitive data beyond the trusted borders of a clinic or physician's office. Fu says that preventing unauthorized re-programming has proven difficult to guarantee. Observers point out, however, that because Fu had the foresight to stress security from the outset, during design, to safeguard cyber trust, patient data and privacy are on track to being secure. As Fu explained last year, "With medical devices, we don't have the luxury to fix security after the fact. This is where our research comes in."

Jason Pontin, editor-in-chief and publisher of *Technology Review,* said, "The TR35 honors young innovators for accomplishments that are poised to have a dramatic impact on the world as we know it. We celebrate their success and look forward to their continued advancement of technology in their respective fields."

Fu and the other TR35 winners for 2009 were featured in the September/October issue of *Technology Review* magazine and were honored in September at the EMTech@MIT 2009 Conference in Cambridge, Massachusetts.

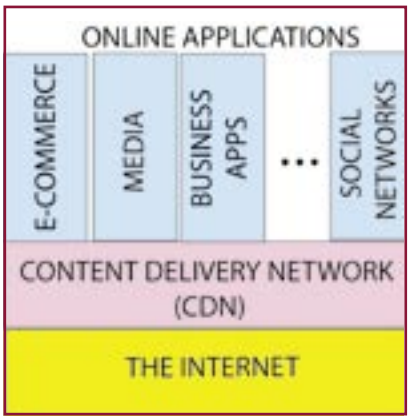
Department Chair Andrew Barto says of Fu's accomplishment, "We are absolutely delighted that Kevin is being honored so visibly for this innovative research. Kevin's focus on the security and privacy issues being raised by pervasive computing not only illustrates his talent for anticipating technological trends, but it also demonstrates his concern for the impact of this technology on our everyday lives. He is a wonderful model of the socially engaged computer scientist!"

SITARAMAN – – – – – – – – *continued from page 1*

actual networks that were unreliable and faulty. "At that time, the focus was networks within large parallel computers. But when the Internet came of its own in the late 1990s, some of the algorithmic paradigms for fault-tolerance and efficient communication seemed serendipitously relevant in the new context," says Sitaraman. Subsequently, he took a leave of absence from the department to join his research colleagues from MIT who had earlier founded Akamai Technologies. At Akamai, he helped architect and build perhaps the world's largest CDN that currently consists of over 50,000 servers in 70 countries, serving hundreds of billions of end-user requests a day, and delivering roughly 20% of the world's Web traffic. Since returning to academia, his research is focused on the foundational and algorithmic challenges in building the CDNs of the future.

A next-generation CDN will likely consist of hundreds of thousands of servers delivering tens of petabits of data per second to hundreds of millions of concurrent users around the world! Building such a large distributed system requires new scientific and algorithmic advances in practically every field of computer science, including distributed protocols for failure recovery, server allocation, and load balancing; techniques for managing large server farms and network deployments; real-time monitoring, alerting, and fault diagnosis; communication protocols for route optimization at Internet scales; massive data aggregation, storage, billing, and analytics; and processes for developing, testing, and releasing software in a fail-safe fashion without downtime.

Consider the futuristic problem of simultaneously transporting a hundred thousand live media shows on the Internet from their respective sources to hundreds of millions of viewers around the world, with very high fidelity. The millions of "optimal" routes used to transport the media streams must avoid failures and congestion hotspots that are inherent in the Internet, and minimize costs as well. Since the pattern of failures and hotspots change frequently, it is essential to efficiently compute and recompute the routes every several minutes. Sitaraman's work models this problem as a large flow optimization where multiple commodities (i.e., streams) must be simultaneously routed on the Internet to meet end-to-end quality guarantees for each viewer watching each stream. While finding the optimal routes exactly is intractable, his work with his colleagues shows that one can use novel techniques that round fractional solutions of linear programs to efficiently produce a provably near-optimal set of routes.



A CDN is a virtual network built over the existing Internet to provide high levels of end-to-end reliability, performance, and scalability for online applications.

"There is a significant amount of academic research on performance, but less in known about cost tradeoffs that are important in practice. Incorporating cost criteria often leads to new and more realistic algorithms for even traditional problems," says Sitaraman. The operating costs of a CDN include both server costs (hardware, power, and colocation) and bandwidth costs (aggregated across thousands of network contracts). An approach to reducing server costs is using Flash-based Solid State Drives (SSDs) that reduce power and cooling costs by orders of magnitude over traditional hard disk drives. But, SSDs function very differently from hard drives, requiring a complete rethinking of key algorithms and data structures. Working with departmental colleagues, Sitaraman has helped devise a new indexing data structure for SSDs called Lazy-Adaptive (LA) Trees that has provable optimality and is an order of magnitude faster than the best traditional solutions such as B+ trees. Sitaraman has also recently studied online algorithms for global server allocation that provably optimize bandwidth costs, when networks charge the CDN for bandwidth using industry-standard metrics such as the 95th-percentile or the average of the traffic usage for the month.

"Inventing the future Internet is a rich source of deep scientific and algorithmic challenges that will continue to inspire my work," says Sitaraman.

Lesser appointed Distinguished Professor

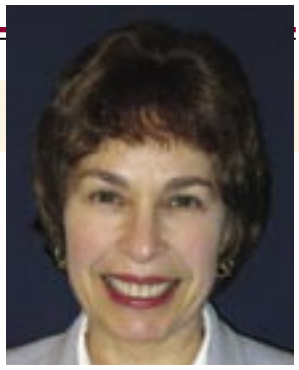
Victor Lesser has been appointed Distinguished Professor by UMass President Jack M. Wilson following the Board of Trustees' approval on June 10.

Provost Charlena Seymour said, "Professor Lesser has clearly established himself as a world-renowned leader in the field of artificial intelligence" and "the founder and one of the most influential researchers in multi-agent systems, a field that provides a powerful model for computing in the 21st century."

According to department chair Andrew Barto, Lesser made path-breaking research contributions to problem-solving architectures, multi-agent systems, real-time AI, and signal understanding. His pioneering work in multi-agent systems helped build a thriving research community made up of thousands of researchers. Lesser was the first to articulate the critical computational research issues in the field and to propose solutions for many of these issues, says Barto.

Earlier this year, Lesser was honored by the International Joint Commission on Artificial Intelligence with its 2009 Award for Research Excellence in recognition of his high-quality research.

Clarke commended for her service to CS



At the College Convocation in September, Professor Lori Clarke received a College of Natural Sciences (CNS) Outstanding Faculty Award in the area of Service.

Clarke has an extensive record of service throughout her 34-year career at UMass Amherst. She has taken on numerous projects to serve the department, the campus and University, the software engineering and computer science research communities, and also to support diversity in computing. Among other activities, she has been a leader, both locally and nationally, in efforts to attract, mentor, and retain women and minorities in Computer Science. "Lori's service contributions are continuing unabated," adds Andrew Barto, Department Chair. Within the department, she has initiated and continues to support the CS Women's Group, which provides networking and mentoring opportunities to women in the department. She participated in activities in the Colleges of Engineering and CNS that are trying to support students from underrepresented groups.

Clarke was instrumental in helping to guide the growth of the department from a faculty of twelve to its current position with over forty-five tenure-track and research faculty and as one of the top twenty computer science departments in the nation. She served as associate chair from 1981-1985 and has continued to provide leadership. She has often served on University and College committees, the Committee on the Status of Women, and the Research Council.

Clarke has also served the national research community. She has served four elected three-year terms on the board of directors of the Computing Research Association (CRA). She recently completed a four-year term as Vice Chair of this organization. Since 2001, Clarke has served on the CRA's subcommittee on the Status of Women in Computing Research (CRA-W), having recently completed a three-year term as co-chair. CRA-W runs over a dozen programs a year to help attract and retain women in computing research. CRA-W received the 2005 National Science Board Award for Public Service and the 2003 Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring for its many mentoring activities. Clarke has served as co-chair of the Grad Cohort workshop, a two-day mentoring workshop for women in the first three years of their graduate program in computing. This year almost three hundred women from 110 institutions participated in this event.

Clarke has also worked tirelessly to support the software engineering research community. The Association for Computing Machinery (ACM) Special Interest Group on Software Engineering (SIGSOFT) is one of the leading professional societies in Software Engineering. Clarke was an elected member of the executive committee from 1985 to 2001, serving as secretary/treasurer, vice chair, chair, and past chair. She has also served on NSF advisory boards.

Grupen to give Distinguished Faculty Lecture

As part of the UMass Amherst Campus Distinguished Faculty Lecture Series, Professor Rod Grupen will speak on "Programming Robots to live Among Us: A Developmental Approach." He will give his talk on March 1, 2010. For more than 30 years, the campus has recognized the distinguished achievements of faculty through the series. Grupen will receive a Chancellor's Medal following his lecture. The Chancellor's Medal is the highest honor bestowed on individuals for exemplary and extraordinary service to the campus.

Rod Grupen, *left*; Rick Adrion, *right*

Adrion named delegate to Board of Trustees

Professor Rick Adrion was named the Faculty Delegate to the University of Massachusetts Board of Trustees. Adrion was chosen in May by a ballot election by the Faculty. He replaces Professor W. Brian O'Connor who held the position for the past 12 years. As Faculty Delegate, Adrion attends all of the Board and Board subcommittee meetings. He also serves on the Intercampus Faculty Council and the Faculty Senate Rules Committee.

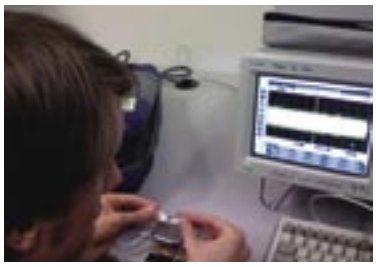
Adrion is the Director of the National Science Foundation (NSF) Commonwealth Alliance for Information Technology Education (CAITE) that designs and carries out comprehensive programs that address under-representation in information technology (IT) education and the workforce. CAITE focuses on women and minorities in groups that are underrepresented in the Massachusetts innovation economy; that is, economically, academically, and socially disadvantaged residents. He is also the Co-Director of the Commonwealth Information Technology Initiative (CITI), Principal Investigator of an NSF Research Experiences for Undergraduates site, and Co-Director of the RIPPLES (Research in Presentation Production for Learning Electronically) multimedia teaching and learning group.

Latest CS research: from mining books to solving crimes

The department's faculty are embarking on many exiting new projects. A few are highlighted below.

Mining a Million Scanned Books - In this NSF-funded project, "Mining a Million Scanned Books: Linguistic and Structure Analysis, Fast Expanded Search, and Improved Optical Character Recognition (OCR)," the UMass Amherst Center for Intelligent Information Retrieval (CIIR), the Perseus Digital Library Project at Tufts University, and the Internet Archive are investigating large-scale information extraction and retrieval technologies for digitized book collections. This research is being carried out using a collection of over a million scanned books that includes 8.5 terabytes of text and half a petabyte of scanned images. Professors James Allan, R. Manmatha, and David Smith are leading the research within the CIIR to develop new approaches to processing the large collection.

Pacemaker Interference - Assistant Professor Kevin Fu and graduate student Ben Ransford, along with researchers from the University of Washington and the Department of Medicine at Beth Israel Deaconess Medical Center, warn about adverse magnetic interactions between headphones and pacemakers in a study published in October in the *HeartRhythm Journal*, the leading specialty journal in cardiology with the largest circulation and readership base. Their research showed that there is clinically significant magnetic interference of implanted cardiac devices by portable headphones. Patients with these devices are now advised to keep the headphones away from their device to avoid magnetic interference.



Ben Ransford, a Computer Science graduate student from UMass Amherst, determines the level of interference on an implantable defibrillator from neodymium magnets inside earbuds.

Advances in OCR - Assistant Professor Erik Learned-Miller (PI) and Professor Andrew McCallum received an NSF grant for their project "Coordinating Language Modeling, Computer Vision, and Machine Learning for Dramatic Advances in Optical Character Recognition." In this project, they are investigating "iterative contextual modeling," an approach to OCR in which high confidence recognitions of easier document portions are used to help in developing document specific models. These models can be related to appearance–for example a sample of correct words can be used to develop a model for the font in a particular document. In addition, the models can be based on language and vocabulary information.

Human Emotion Sensors - Research Professor Beverly Woolf continues her research on sensors that detect human emotion. Woolf and her research team have demonstrated that intelligent tutoring systems can provide adaptive feedback based on an individual student's affective state. Their primary research goal is to identify whether a dependency exists between students' reported emotions and their learning, motivation, and attitudes toward a subject. The sensors are placed on a student's chair, mouse, monitor, and wrist to provide data about posture, movement, grip tension, facially expressed mental states and arousal. Woolf's researchers recently tested the system with 600 students in Deerfield, MA to confirm that their sensors can predict student emotion (frustrated, bored, etc.) with up to 80% accuracy in comparison with the student's statement of his/her emotion.

Molecular Playground - The molecular aspects of nature are too often viewed as inaccessible and uninteresting to the general public. While the public can appreciate the beauty of a flower or a swan, the molecular basis of these organisms goes unnoticed. Professor Emeritus Allen Hanson and colleagues from chemistry, microbiology, and computer science are working on a project to get the organisms noticed. They are developing a system for displaying large-scale interactive molecules in prominent public spaces. The first such system has been installed in the new Integrated Sciences Building on campus. The aim is to capture the public's attention and to prod individuals to explore personally a vast array of molecular structures in a human-size "molecular playground." The local Playground installation consists of a projector, an infra-red (IR) illuminator, and a camera fitted with a filter that blocks visible light, but passes IR. In this way, the camera "sees" the person playing with the image, but does not see the projected image itself. The camera tracks movement and the software then decides what to use as a trigger to tell the system to stop the pre-programmed animation and instead deliver "rotate" commands to the system.

Forensic Analysis - Associate Professor Brian Levine (PI) is working on an NSF-funded project to significantly advance forensic analysis for crimes involving mobile systems. While current methods and legislation focus heavily on logical identifiers, Levine and UMass Amherst Electrical & Computer Engineering Professor Dennis Goeckel (co-PI) are designing, evaluating, and deploying new forensic techniques that focus on consistent and trackable characteristics of mobile computing. Additionally, their work plays an important role in understanding the limits of personal privacy in these settings. They are developing new radio fingerprinting techniques that detect identifying information present in a radio's low-level components, and the team is developing novel techniques of traffic analysis that determine the source of encrypted Web traffic. Their research will directly assist law enforcement that investigate network trafficking of images of child sexual exploitation, demonstrating the usability of trustworthy computing.

CS Department to be important partner in high-performance computing center planned for Holyoke

In October, Massachusetts Governor Deval Patrick announced that groundbreaking of Holyoke's High Performance Computing Center (HPCC) is on track to begin next fall, with a target completion date set for late 2011. Department faculty involved in the project planning include Distinguished Professor Jim Kurose and Professors Rick Adrion and Prashant Shenoy.

The public-private collaboration, which recently completed a successful 120-day planning process, will establish Massachusetts as a global leader in the development and application of the next generation of computing technologies. Partners in this collaboration include the University of Massachusetts, MIT, Boston University, EMC, Cisco, Accenture, the Pioneer Regional Valley Planning Council (PVPC), and the City of Holyoke.

Initially announced by the Governor and the collaborative partners on June 11, 2009, the project will create a world-class, green, high performance computing center in Holyoke that will provide an infrastructure for research computing and a collaborative research agenda in advanced computing and applications such as life sciences, clean energy, and green computing.

"Holyoke is absolutely the right place to locate this project. It has clean, green, and cost-effective hydroelectric energy, world-class industry partners, a broad spectrum of outstanding educational institutions, and access to a talented workforce both locally and across the Commonwealth," says Distinguished Professor Jim Kurose, Executive Dean of the College of Natural Sciences, who helped envision the center.

The HPCC will enable the region's internationally recognized academic institutions to establish leading-edge collaborative research programs and collaborations with local community colleges and high schools that will include hands-on education, internships and outreach to students. On the research front, the HPCC will facilitate collaboration among stakeholders in the areas of green computing (both the "greening of computing" - the design of energy-efficient hardware, software, and systems operations - as well as "computing for greening" - the use of computing to make organizations, businesses, homes and physical infrastructure more carbon-efficient) and in computationally-intensive research in areas including life sciences and human health, health care delivery, energy and the environment. The department and the Commonwealth Alliance for Information Technology Education (CAITE) are working with Springfield Technical Community College and Holyoke Community College to develop new curricula, events, training programs, internships, and outreach activities for the new center. Kurose, Shenoy, and Adrion are working closely with the Massachusetts Department of Housing and Economic Development, the Adams Institute, and the PVPC to develop the vision for a Holyoke Innovation District that would grow up around the HPCC.

High-performance computing centers use powerful, networked or special-purpose computers and computer clusters to handle massive amounts of data and tackle complex problems, Kurose noted. They demand high power and cooling capacity, both of which are available at lower cost from Holyoke's environmentally friendly hydroelectric facilities. Kurose explained that a few years ago the cost of housing and maintaining large numbers of computers drew even with, then significantly exceeded, the cost of the computers themselves, and that gap continues to widen. "If energy is ultimately going to drive the costs," he noted, "then you look around for who can deliver that."

Ben Barnhart photo

Jim Kurose

Now, using the same water source that powered the local industrial revolution 150 years ago, the HPCC project can be a model of energy efficiency for other such centers. Holyoke also offers low real estate costs and access to high-speed fiber optic data lines along I-91 and the Massachusetts Turnpike.

Kurose said the Holyoke HPCC idea evolved from a meeting last year hosted by MIT President Susan Hockfield that included Governor Patrick, UMass President Jack Wilson and other university leaders, along with leaders from the information technology industry in Massachusetts. The project website, **InnovateHolyoke.com**, is the official destination for the latest information related to the center.

CIFellowship recipients

Jeffrey Johns, Victoria Manfredi, and Leena Razzaq have been named 2009 Computing Innovation Fellows (CIFellows). Johns and Manfredi are recent UMass Amherst Computer Science Ph.D. graduates, and Leena Reeza is a Postdoctoral Researcher doing her Fellowship in the department.

Johns will begin his fellowship work at Duke University under the mentorship of Professor Ronald Parr. Manfredi began her fellowship work at Boston University under the mentorship of Professor Mark Crovella. Razzaq, who received her Ph.D. in Computer Science from Worcester Polytechnic University, is being mentored by Research Professor Beverly Woolf.

The CIFellows Project is sponsored by The Computing Community Consortium (CCC) and the Computing Research Association (CRA), with funding from the National Science Foundation, and allows new Ph.D. graduates to obtain one-to-two year postdoctoral positions at host organizations including universities, industrial research laboratories, and other organizations that advance the field of computing and its positive impact on society.

University of Massachusetts Amherst
Fall / Winter 2009

Alum Matters

A newsletter for alums of the Department of Computer Science

Bryant York: a retrospective

As I near the end of my useful lifetime as a slowly moving part in the machinery of life, I am afforded the luxury of waxing philosophical. Young people, gird yourselves! I remember a time when there were almost no African American and very few women computer scientists. The world was a very different place and American society was experiencing severe diversity growing pains. The civil rights movement of the 1960s (including the accompanying turmoil); the assassinations of John F. Kennedy, Robert Kennedy, Martin Luther King, Jr., and Malcolm X; the Vietnam War; the quest to put a man on the moon; the women's liberation movement; and the birth of the Information Era, were only part of the social context in which many early computer science departments were conceived and born. These were also the formative years of my life. The UMASS Amherst Computer Science Department (formerly COINS) was founded during this time, rapidly coming of age in the 1970s, defining itself and establishing a departmental culture that engenders success in its graduates. In order to understand why COINS was such a good fit for me, you need to understand a little about me and a little about COINS.

As a young child growing up in the segregated, federally-subsidized housing projects of Boston in the 1950s, I thought there were only two kinds of people – Black people and White people. I spent countless hours devising ways to avoid white people, as overt racism was a simple fact of daily life. Our annual race riots lasted approximately three weeks each summer, starting on the Fourth of July and dissipating by the end of the month. During this period my family would often sleep in the windowless hallway of our apartment in order to avoid the nightly rocks, beer bottles and Molotov cocktails coming through the windows. Fortunately, I had three things going for me: (1) My father was a strict disciplinarian who believed that any obstacle could be overcome with hard work; (2) My mother was a hard-working pragmatist who epitomized human compassion; and (3) I was a middle child. Both parents were raised in Catholic orphanages and reached adulthood during the Great Depression. Favorite mixed metaphors and quotes (repeatable in public) from my mother include, "It is better to be lucky than smart," "Work hard, keep your head down, and be invisible," "Don't let your light skin get in the way of your blackness," "Maintain your focus while others are trying to distract you," "There, but for the grace of God goes me!" and, "Please consider the priesthood."

In the seventh grade, I was sent to the Boston Latin School where I was one of a handful of blacks. During these years, I learned the lost art of managing two distinct identities – Bryant York, student in the white setting at Boston Latin School and Bryant York, black boy in Orchard Park Housing Project. During my first year at Latin School, Sputnik went up and fear overtook the nation. I was selected for an advanced mathematics program and sent off to MIT every summer. It was at Latin School that I first realized there were many different kinds of white people (English Americans, Italian Americans, German Americans, Polish Americans …) and where I first learned of anti-Semitism, anti-feminism, and the diversity and gradations of bigotries. I graduated Latin School in the early 1960s; majored in mathematics at Brandeis University where I learned FORTRAN; was invited to the IBM Cambridge Scientific Center where I learned LISP, and APL; subsequently worked for two years; earned an M.S. in management at MIT's Sloan School; and worked three more years before applying to the computer science program at UMass Amherst. What prompted me to apply to UMass? I was sitting in a book store in 1974 when I came across Michael Arbib's book, *Brains, Machines and Mathematics,* for me a serendipitous, life-changing experience. I had always loved mathematics and computers so I hopped into my car, drove to UMASS Amherst, and without an appointment went into to Dr. Arbib's office. He spoke with me for over an hour and encouraged me to apply to COINS. In the fall of 1974 at the advanced age of 29, I quit my high-paying job and joined the COINS community.

COINS was a vibrant research community in which student autonomy was highly valued. The faculty at that time included Michael Arbib, Bill Kilmer, Cax Foster, Nico Spinelli, Ed Riseman, Robbie Moll, Dan Fishman, and Bob Taylor. At the time Al Hanson was still at Hamp-

shire College while Connie Wogrin and Dave Stemple had taken responsibility for firing up our new CDC 6600 in the computer center. Lori Clarke, Vic Lesser, Jack Wileden and Bob Graham joined the faculty during the next two years. There was a remarkably low level of competition and an extraordinarily high level of cooperation among the early cohorts of graduate students, which included Art Karshmer, Elliot Soloway, Ed Fisher, Jim Stanley, Peter Burt, Fanya Montalvo, Charlie Welty, Paul Nagin, Debra Richardson, Tom Williams, Andrew Singer, John Woods, Len Wesley, John Lowrance, Mitch Zolliker, Neal Ogden, Kurt Konolige, Bev Woolf, Daryl Lawton, Dan Corkill, Ralf Kohler and others. It was at COINS that I first experienced white friendship, the freedom to express my ideas, no need to be invisible, no need to keep my head down, and no need to maintain dual identities. I developed a number of lifelong friendships that I cherish daily.

COINS presented a broad array of enrichment opportunities. Here is a partial list of the kinds of positive experiences from that time. Ed Riseman and Michael Arbib invited David Marr, one of the leading researchers in the field of computer vision, to UMass Amherst and took me to lunch with him to discuss my dissertation research in computer vision. David Huffman and Takeo Kanade were invited to give presentations at UMass Amherst followed by informal conversations with graduate students. Each of them encouraged my work and provided me with sets of origami figures that were relevant to my research. In 1975 I had the opportunity to teach a course inside the Pittsfield State Prison through the UMass Amherst University Without Walls Program. It was also the year in which I had my first exposure to the Moore Method of teaching in a course taught by Robbie Moll. Ed Riseman and Al Hanson invited a number of celebrated researchers in computer vision (including Nils Nilsson, Ruzena Bajcsy, Azriel Rosenfeld, Raj Reddy, and Marty Tenenbaum) to a poster session in order to provide critical commentary on the work of their students. In 1977, Ed and Al took the VISIONS team to the first International Joint Conferences in Artificial Intelligence (IJCAI) and introduced us to the leading researchers in the field. All of the people mentioned above have played significant roles in my life and career and I am truly grateful for these opportunities gated by my presence in the COINS community. Finally, while I was in the writing stage of my dissertation, my mother called my siblings and me to Boston to announce that she had terminal lung cancer and about three months to live. My mother insisted that I return to UMass Amherst and complete the dissertation. Being an almost-dutiful son (I had eschewed the priesthood), I acceded to her request. Back at UMassAmherst, thinking and writing became impossible and Ed detected that something was amiss. By this time Ed and I had become old buddies. We lived in the same apartment complex and often commuted to school together. He called me to his office and got me to open up. Once he knew my problem, he insisted that I go home to care for my dying mother. He pointed out that this was one of the few situations in which it is permissible to ignore a mother's orders. I'll never forget Ed Riseman!

Recently many people have begun to ask me, "With the kind of research background and opportunities afforded by COINS, how did you end up as a doorman?" My reply is usually something like, "For a kid who was born with a plastic spoon in his mouth, I didn't do so badly."

If someone had told me forty years ago that I would spend most of my professional life as a doorman, I would have been puzzled and laughed, but that is what has happened. Service has clearly dominated my professional career. Although I enjoy research and teaching immensely, they have always taken a back seat. I could have kept my head down, remained somewhat invisible, and tried to join the

continued on page 10

York in 1957

In September, CIIR alum **Ben Carterette** (Ph.D. '08) received the Best Paper Award at the 2nd International Conference on the Theory of Information Retrieval (ICTIR '09) for his paper, "An Analysis of NP-Completeness in Novelty and Diversity Ranking." The award was sponsored by Yahoo! Carterette is currently as Assistant Professor at the University of Delaware.

This summer, **Anita Raja** (Ph.D. '03) was promoted to Associate Professor with tenure in the Department of Software and Information Systems at the University of North Carolina at Charlotte. Dr. Raja was also married in June to Dr. Cephas Swamidoss. Their wedding was held in Chennai, India.

In July, Microsoft appointed **Steven Sinofsky** (M.S. '89) as President of its Windows and Windows Live Division. Prior to this position, Sinofsky was Senior Vice President of the Windows and Windows Live engineering group. He previously oversaw the development of the Microsoft Office system of programs, servers and services.

Evan Smith (M.S. '91, BS '88) is currently in his fourth year of medical school at the University of Chicago. Prior to returning to school, Smith was the Director of Engineering at Wellogic.

SAVE THE DATE: Alum Banquet 2010

The second Outstanding Achievement and Advocacy (OAA) Awards Banquet will be held on campus on the evening of Friday, April 30, 2010. The OAA awards program recognizes the achievement of our alums in such areas as entrepreneurship, scientific research, and education. Please join us to celebrate the accomplishments of our award recipients and to socialize with faculty and fellow alums. Details on the events can be found at **www.cs.umass.edu/oaa2010.**

YORK – – – – – – – – – – *continued from page 9*

100+ publications club, *possibly* earning the respect of some of my white colleagues. However, I saw an important opportunity, a chance to change the culture of the field of computer science. The job description read, *"Keep the door to computer science propped open long enough for minorities and women and persons with disabilities to enter in sufficient numbers that they are able to participate in a normal, unfettered manner without distractions."* One of my great rewards is that now I can attend a conference in almost any sub-discipline of computer science and see at least one other black face. Currently there are over 170 black PhDs in computer science (as well as over 200 black graduate students in the pipeline) doing research and teaching in academic, governmental, and industrial institutions around the world. Would these numbers have resulted without major effort by me and many others? It's highly unlikely. I can point to several young researchers, whom I have mentored, who are well on their way to membership in the 100+ publications club. So, was it worth being tokenized? Was it worth serving on all those committees and advisory boards? I reply, "Definitely yes!" My legacy is that I helped to create the luxury of the single-identity black computer science researcher. In 2006, when I was named an ACM Fellow, there were a number of negative reactions, and I overheard one of my white colleagues say, *"If Bryant can become an ACM Fellow, then anyone can."* The implication was that the ACM had lowered its standards for Fellow in order to admit me. Initially I was hurt, and then I was heartened. The ACM had made a statement about its values. It valued contributions to the overall health of the computing discipline as well as excellence in research.

So now you know a little about me and a little about UMass Amherst Computer Science (COINS).

Now, I offer my final pontification. In your later years, when asked to fit a straight line or even a piecewise linear curve through your life's achievements, you should respectfully decline. Backtracking is a metaphor for life. One should never be afraid to backup and start down a different path. Success is measured in many different ways. What's most important is to live a life worth living. Life is a sequence of human amplification opportunities and you must take advantage of these opportunities in order to be in a position to create such opportunities for others. Now the world is flat and the United States has a black president. It couldn't be more different from the world in which I grew up 50 years ago. What of value can a modern student take from this exposition? What can an old black man, who never had a white friend until he was 30 years old, tell a young black researcher (who has had white friends since birth) about how the modern world works? Regardless of your achievements, a person's character must transcend the kinds of situations in which you find yourself. On a national/international level, I was fortunate to live during the times of such icons as Nelson Mandela, Robert Kennedy, Martin Luther King, Jr., John F. Kennedy, Muhammad Ali, and Jackie Robinson. What I learned at UMass Amherst/COINS was how to recognize and admire those heroic values in my everyday colleagues.

Dr. York is currently a Professor of Computer Science at Portland State University and Co-Director of the Laboratory for Learning and Adaptive Systems. He received an M.S. and Ph.D. in Computer Science from UMass Amherst in 1976 and 1981, respectively.

CIIR Reunion

The Center for Intelligent Information Retrieval (CIIR) held a reunion during the 32nd Annual International ACM SIGIR (Special Interest Group on Information Retrieval) Conference held in Boston, MA in July. In addition to a social gathering of CIIR alums and associates who had traveled from around the world for SIGIR 2009, the group celebrated Bruce Croft's 30th year at UMass Amherst and James Allan's 15th year in the CIIR. Allan was General Co-Chair of SIGIR 2009, a five-day conference with more than 500 attendees.

Front Row (left to right): Van Dang, Yun Zhou, Dave Lewis, Carol Broverman, Bruce Croft, Howard Turtle, Jamie Callan, Jean Joyce, Matt Lease. 2nd row: Bob Krovetz, Jeremy Pickens, Kate Moruzzi, Hema Raghavan, Kyung Soon Lee, Elif Aktolga, Vanessa Murdock, Sai (Chandu) Ravela, Jiwoon Jeon, Don Metzler; 3rd row: Xing Yi, Fernando Diaz, Chirag Shah, Xiaobing Xue, Sudheer Gaddam, Jangwon Seo, Jinxi Xu, Hongmin Shu, R. Manmatha, Eric Brown, Trevor Strohman; back row: Jinyoung Kim, Ben Carterette, Anton Leuski, Victor Lavrenko, James Allan, Mark Smucker, Dirk Mahling, David Harper, and Mark Sanderson.

Department welcomes three adjunct faculty

This fall, Jeannie Albrecht and Michael Zink joined the Department as Adjunct Assistant Professors and Jane E. Fountain joined as an Adjunct Professor.

Dr. Albrecht is an Assistant Professor of Computer Science at Williams College. Her research interests include the design and performance of distributed systems and computer networks, particularly focusing on reliability, scalability, and extensibility achieved over the wide-area. Albrecht is currently working on the Global Environment for Network Innovations (GENI) project sponsored by the National Science Foundation. She is Co-chair for the GENI Working Group on Experiment Workflow and Services. In 2009, she received an NSF CAREER award for her project "Mobile Application Management." Albrecht received a Ph.D. in Computer Science from the University of California San Diego in 2007, an M.S. in Computer Science from Duke University in 2003, and a B.S. in Math and Computer Science from Gettysburg College in 2001.

Dr. Fountain is a Professor of Political Science and Public Policy at UMass Amherst. Previously, she served for 16 years on the faculty of the John F. Kennedy School of Government at Harvard University. Fountain is the founder and Director of the National Center for Digital Government, based at UMass Amherst, which was established with support from the National Science Foundation to build research and infrastructure for the emerging field of research on technology and governance. She also directs the new Science, Technology, and Society Initiative, a campus-wide effort at UMass Amherst, that serves as a catalyst for research partnerships between social, natural, and physical scientists on campus and beyond. Fountain is currently a member of the World Economic Forum Global Advisory Council on the Future of Government, and a member of American Bar Association blue ribbon commission on the Future of e-Rulemaking. She has served on several advisory bodies for organizations including the Social Science Research Council, the Internet Policy Institute, and the National Science Foundation. Fountain holds a Ph.D. from Yale University, in organizational behavior and in political science, and graduate degrees from Harvard and Yale. She has been a Radcliffe Fellow, a Yale Fellow, and a Mellon Fellow.

Dr. Zink is an Assistant Professor of Electrical and Computer Engineering (ECE) at UMass Amherst, having joined ECE in fall 2009 after spending five years in the UMass Amherst Computer Science Department. Most recently, Zink was an Assistant Research Professor in the CS Department as a member of the leadership team of the UMass Amherst -led NSF Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (CASA). Within ECE, Zink is affiliated with the Microwave Remote Sensing Laboratory. He works in the fields of sensor and distribution networks for high bandwidth data. Further research interests are in wide-area multimedia distribution for wired and wireless environments and network protocols. He is one of the developers of the KOMSSYS streaming platform. Zink is a member of IEEE and ACM. He received the Best Paper Award at Multimedia Computing and Networking in 2008. Zink received his M.S. (Diploma) in Electrical Engineering and Information Technology and his Ph.D (Dr.-

continued on page 15

Massachusetts research labs host CS REU students

In July, program staff, faculty, and nineteen students in the Department's National Science Foundation Research Experience for Undergraduates (REU) summer program spent the day in Cambridge at Google Labs, the IBM Center for Social Software, and Microsoft New England Research & Development Center. The group received a tour of each facility and was able to discuss projects and career options at each stop. During the visit to Google Cambridge, Site Director Steve Vinter (UMass Amherst CS PhD '85) gave a one-hour talk to the students followed by a question and answer session and lunch. Many of the REU post-program surveys indicated that students felt this field trip was very valuable in exposing them to the world of research outside of academia and a highlight of their summer research activities, says Wendy Cooper, REU Program Manager.

Recent Computer Science Ph.D. graduates (AY 2008-2009)

Martin Allen: *Agent Interactions in Decentralized Environments;* (Shlomo Zilberstein, Advisor); Feb. 2009; Mellon Post Graduate Fellow, Connecticut College.

The decentralized partially observable Markov decision process (Dec-POMDP) is a powerful formal model for multiagent problems where cooperative action is optimal, but agents act based on local data alone. Unfortunately, Dec-POMDPs are intractable: NEXP-complete in the worst case, with experiments indicating optimal solutions are infeasible. Research therefore often focuses on special subclasses, restricting the interaction between agents. In some cases, structured interactions provably reduce complexity. Empirical results suggest this also holds where formal proofs are lacking.

I establish novel complexity results for some popular restricted-interaction models, showing worst-case intractability to be more wide-spread than previously known. I therefore propose new information-theoretic ways of analyzing the actual difficulty of Dec-POMDPs. Empirical results show that these measures correlate with problem difficulty, allowing us to apply optimal algorithms more widely, and intelligently tuning performance of approximate methods as well.

Benjamin Carterette: *Low-Cost and Robust Evaluation of Information Retrieval Systems;* (James Allan, Advisor); Sept. 2008; Assistant Professor, University of Delaware.

Research in Information Retrieval has progressed against a background of rapidly increasing corpus size and heterogeneity, with every advance in technology quickly followed by a desire to organize and search more unstructured, more heterogeneous, and even bigger corpora. But as retrieval problems get larger and more complicated, evaluating the ranking performance of a retrieval engine gets harder: evaluation requires human judgments of the relevance of documents to queries, and for very large corpora the cost of acquiring these judgments may be insurmountable. This cost limits the types of problems researchers can study as well as the data they can be studied on.

I present methods for understanding performance differences between retrieval engines in the presence of missing and noisy relevance judgments. The work introduces a model of the cost of experimentation that incorporates the cost of human judgments as well as the cost of drawing incorrect conclusions about differences between engines in both the training and testing phases of engine development. Through adopting a view of evaluation that is more concerned with distributions over performance differences rather than estimates of absolute performance, the expected cost can be minimized so as to reliably differentiate between engines with less than 1% of the human effort that has been used in past experiments without sacrificing the ability to reuse the judgments in future experiments.

Rachel Cobleigh: *PROPEL: An Approach Supporting User Guidance in Developing Precise and Accessible Property Specifications;* (Lori A. Clarke and George S. Avrunin, Advisors); Sept. 2008; Usability Specialist, The MathWorks, Inc.

Property specifications concisely describe an aspect of system behavior. Although each property has a narrow focus, it can be difficult to specify a property correctly. There are subtle, but important, details in desired system behavior that can often be overlooked. Property specifications should be precise enough to support automated analyses and plain enough to be easily understood by system developers. Property specifications can be written in mathematical formalisms, which provide precision, but are difficult to understand. In practice, property specifications are usually written in natural language, but such informality often makes them ambiguous and they cannot be used in many types of automated analyses.

We developed a PROPerty ELucidator tool (PROPEL) which aims to make the job of specifying and understanding properties easier by providing templates that build on commonly-occurring property patterns. These templates indicate variations in the subtle details that must be considered. PROPEL provides three alternative property views: graphical finite-state automata for precision; "disciplined" natural language, for understandability; and question trees, for guidance in selecting a template.

To evaluate PROPEL, we used it to specify properties in five case studies in the medical domain. The results indicate that our approach is promising because it was effective at specifying most of the properties we encountered.

Paul Dickson: *Automatic Capture and Indexing of Lectures; (*W. Richards Adrion and Allen R. Hanson, Advisors); Sept. 2008; Visiting Assistant Professor, Hampshire College.

This dissertation describes the design, implementation, and testing of Presentations Automatically Organized from Lectures (PAOL), a content capture system that automatically captures lectures. PAOL-captured presentations include a video, a thumbnail-based table of contents, and windows containing the material currently being presented on computer and whiteboard. These presentations are searchable using the table of contents and give the viewer access to all material presented in the lecture. The capture process occurs transparently to the lecturer with the only requirement being that the lecturer wear a wireless microphone. No training or installation of special software is required. These presentations are created to give students access to class content after the class and through the table of contents allow students to review specific portions of the material presented.

PAOL is divided into 3 subsystems: video creation, computer capture, and whiteboard capture. Image processing techniques are used to create a video of the lecturer from fixed-view high-resolution cameras focused on the whiteboard. Image processing techniques are also used to identify significant changes on the computer and whiteboard in order to form the table of contents. The whiteboard images are also processed to remove the instructor, heighten contrast, and improve clarity.

Ao Feng: *Incident Threading in News; (*James Allan, Advisor); Sept. 2008; Software Development Engineer, Amazon.

With an overwhelming volume of news reports currently available, there is an increasing need for automatic techniques to analyze and present news to a general reader in a meaningful and efficient manner. We explore *incident threading* as a possible solution to this problem. All text that describes the occurrence of a real-world happening is merged into a news incident, and incidents are organized in a network with dependencies of predefined types.

Earlier attempts at this problem have assumed that a news story covers a single topic. We move beyond that limitation to introduce *passage threading*, which processes news at the passage level. First, we develop a new testbed for this research and extend the evaluation methods to address new granularity issues. Then, a three-stage algorithm is described that identifies on-subject passages, groups them into incidents, and establishes links between related incidents. Finally, we observe significant improvement over earlier work when we optimize the harmonic mean of the appropriate evaluation measures. The resulting

performance exceeds the level that a calibration study shows is necessary to support a reading comprehension task.

Yu Gu: *Scalable Techniques for Network Control and Evaluation; (*Donald F. Towsley, Advisor); Sept. 2008; Research Staff Member, NEC Lab.

The Internet has been experiencing significant changes in its scale, capacity, user population, traffic volume and the variety and quantity of online applications and devices. Its continuing evolution demands scalable technologies that provide sustained support for a network with increasing scale and complexity.

In this thesis, we propose several scalable techniques that apply to network control and evaluation. Specifically, we consider the following three problems: First, what is the appropriate congestion control algorithm for networks as capacities get higher and higher and workloads evolve? Second, how does one efficiently measure one-way packet loss rates along paths in a large network? And third, how does one simulate large high bandwidth IP networks so as to obtain detailed packet level information?

Gary Holness: *A Statistical Approach to Improving Accuracy in Classifier Ensembles; (*Paul E. Utgoff, Advisor); Sept. 2008; Lead Research Scientist, Lockheed Martin Advanced Technology Laboratories.

Popular ensemble classifier induction algorithms, such as bagging and boosting, construct the ensemble by optimizing component classifiers in isolation. The controllable degrees of freedom in an ensemble include the instance selection and feature selection for each component classifier. Zenobi et al demonstrated that ensemble construction should consider a classifier's contribution to ensemble accuracy and diversity even at the expense of individual classifier performance. To tradeoff individual accuracy against ensemble accuracy and diversity, a component classifier inducer requires knowledge of the choices made by the other ensemble members.

We introduce an approach, called DiSCO, that exercises direct control over the tradeoff between diversity and error by sharing ensemble-wide information on instance selection during training. In this work, we explore a method for training the component classifiers collectively by sharing information about training set selection. This allows our algorithm to build ensembles whose component classifiers select complementary error distributions that maximize diversity while minimizing ensemble error directly. Treating ensemble construction as an optimization problem, we explore approaches using local search, global search, and stochastic methods. In ensemble classification research, how to use diversity to build effective classifier teams is an open question. We also provide a method that uses entropy as a measure of diversity to train an ensemble classifier.

Ozgur Simsek: *Behavioral building blocks for autonomous agents: description, identification, and learning; (*Andrew G. Barto, Advisor); Sept. 2008; Postdoctoral Research Fellow, Max Planck Institute, Berlin, Germany.

The broad problem I address in my dissertation is design of autonomous agents that can learn how to achieve desired behaviors in large, complex environments. I focus on one essential design component: the ability to form new behavioral units (or skills) from existing ones. For instance, how can a robot (one that needs to manipulate objects) figure out that grasping is a useful behavior, learn how to perform this behavior, and use it as a behavioral building block in the future? Similarly, how can a tic-tac-toe player go through an analogous process to form a behavior for creating a fork on the board?

In addressing this problem, the main assumption I make is that the agent's interaction with its environment forms a Markov decision process. I make three primary contributions. First, I propose a concrete definition of what makes a useful skill. This definition uses a graphical representation of the agent's interaction with its environment and is based on an analysis of shortest paths on this graph. Second, I introduce several online algorithms for identifying such skills. And third, I formulate the exploration problem (of how to efficiently acquire a desired skill) as an optimization problem and introduce a reinforcement learning algorithm that solves it approximately.

Mark Smucker: *Evaluation of Find-Similar with Simulation and Network Analysis; (*James Allan, Advisor); Sept. 2008; Assistant Professor, University of Waterloo.

In this dissertation, we show how the addition of a simple user interaction mechanism, find-similar, can improve information retrieval (IR) quality by making it easier for users to navigate from relevant documents to other relevant documents. Find-similar allows a user to request documents similar to a given document. In the first part of the dissertation, we measure find-similar's retrieval potential through simulation of a user's behavior with hypothetical user interfaces. We show that find-similar has the potential to improve the retrieval quality of a state-of-the-art IR system by 23% and match the performance of relevance feedback. We also show how find-similar responds to varying initial conditions and acts to compensate for poor retrieval quality. In the second part of the dissertation, we characterize find-similar by measuring the quality of the document networks formed by find-similar's document-to-document similarity measure. Find-similar effectively creates links between documents that allow the user to navigate documents by similarity. We show that find-similar's similarity measure affects the navigability of the document network and how a query-biased similarity measure can improve find-similar. We develop measures of network navigability and show that find-similar should make the World Wide Web more navigable. Taken together, the simulation of find-similar and the measurement of the navigability of document networks shows how find-similar as a simple user interaction mechanism can improve a user's ability to find relevant documents.

Xuerui Wang: *Structured Topic Models: Jointly Modeling Words and Their Accompanying Modalities; (*Andrew McCallum, Advisor); May, 2009; Scientist, Yahoo! Labs.

The abundance of textual data in the information age poses an immense challenge for us: how to perform large-scale inference to understand and utilize this overwhelming amount of information. We develop effective and efficient statistical topic models for massive text collections by taking care of extra information from other modalities in addition to the text itself. Text documents are not just text; various additional information is naturally interleaved with text. Most of the previous work, however, pays attention to only one modality at a time and ignores the others. I present a series of probabilistic topic models to show how we can bridge multiple modalities of information, in a united fashion, for various tasks. Interestingly, joint inference over multiple modalities leads to many findings that cannot be discovered from just one modality alone, which are clear evidence that we can better un-

continued on page 15

Faculty News





In promotions effective in September, **Victor Lesser** was promoted to Distinguished Professor, **Andrew McCallum** and **Prashant Shenoy** were promoted to full Professor, **Mark Corner** was promoted to Associate Professor, and **David Jensen** received tenure.

At the November 7 UMass Amherst football game, Chancellor Robert Holub recognized Assistant Professor **Kevin Fu** for his achievements. Fu was on-field at McGuirk stadium during a pause in the game for the recognition.



Professors **Rod Grupen** (20 years), **Robbie Moll** (35 years), and **Jack Wileden** (30 years) each received Length of Service Awards for their years of service on campus.

At the 2009 International Conference on Artificial Intelligence in Education, Research Professor **Beverly Woolf,** Research Scientist **Ivon Arroyo,** graduate student **David Cooper**, and Arizona State University co-authors received the Best Paper Award for "Emotion Sensors go to School," and they also received the Best Interactive Event Award for their Demonstration.

Professor **Prashant Shenoy** received a 2009 IBM Faculty Award. This worldwide competitive program was established to foster collaboration between university faculty members and those in IBM research, development, and services organizations.

Associate Professor **Emery Berger** is co-chair of the 2010 Virtual Execution Environments (VEE) Conference. Berger gave an invited talk at Cornell University this fall.

Distinguished Professor **Don Towsley** gave three keynote addresses recently: at the Information Theory Workshop on Networking and Information Theory held in Volos, Greece; at the International Conference on Wireless Algorithms, Systems and Applications held in Boston, MA; and at the 17th Annual Meeting of the IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems held in London, UK.

Distinguished University Professor Emeritus **Arnold Rosenberg** is spending part of his retirement as a research professor at Colorado State University. He and Colorado colleagues recently received an NSF grant to design more robust and dependable computer systems.

Along with University of Toronto colleagues, Professor **Shlomo Zilberstein** and graduate student **Siddharth Srivastava** co-organized a workshop on Generalized Planning as a part of the 19th International Conference on Automated Planning and Scheduling (ICAPS 2009).

Adjunct Professor **Ileana Streinu** was named by the Five Colleges as one of three new 40th Anniversary Professors. The program was created in 2005 to give veteran faculty teaching opportunities on campuses other than their own. Streinu, a Smith College professor, is creating a multi-institution research group of graduate and undergraduate students.

Steve Constantine was promoted to Senior Lecturer and CS alum **Paul Dickson** (Ph.D. '08) was named an Adjunct Lecturer in the department.

Researcher News

Parviz Kermaqni joined the Laboratory for Advanced Systems Software as a Senior Research Fellow.

Leena Razzaq joined the Center for Knowledge Communication as a Postdoctoral Research Associate.

Ahad Mammadov is a Visiting Scholar from Qafqaz University, Azerbaijan working with Professor Andrew Barto.

Barb Staudt Lerner, Associate Professor at Mount Holyoke College, is a Visiting Research Scholar with the Laboratory for Advanced Software Engineering Research (LASER).

Guillaume Viguier-Just joined LASER as a Research Fellow.

Student News

Graduate students **Michael Hay** and **Chao Li,** along with co-authors **Gerome Miklau** and **David Jensen**, received the Best Student Paper Award at the 2009 International Conference on Data Mining (ICDM) for their paper "Accurate Estimation of the Degree Distribution of Private Networks."

Graduate student **Daniel Menasche**, along with Professors **Don Towsley** and **Arun Venkataramani,** received the Best Paper Award at the ACM International Conference on emerging Networking EXperiments and Technologies (CoNEXT 2009) for "Content Availability in Swarming Systems: Models, Measurement and Bundling Implications." The paper was also co-authored by Antonio Rocha (U. Fed. Rio de Janeiro) and Bin Li (Tsinghua U.).

Moss receives Chancellor's Award

During the fifth annual UMass Amherst Faculty Convocation in October, Chancellor Robert Holub presented Professor Eliot Moss with the Chancellor's Award for Outstanding Accomplishments in Research and Creative Activity. He was recognized for his leadership in founding the research community on transactional computing and revitalizing the area of storage management (otherwise known as garbage collection). Moss is one of seven nationally known faculty members to be presented with the award this year.

Undergraduate study room opened

This fall, the department opened an undergraduate study space in the Computer Science building, U-Space2, to give students a place to congregate, study, and relax. The room was unveiled at the undergraduate *First Friday* event held during the first week of classes.

The space became available in the CS building when some of the department's robotics activities consolidated in the Lederle Graduate Research Center (LGRC). The first U-Space for undergraduates is also housed in the LGRC. With the renovating of U-Space2, the students now have a place to study in both buildings.

U-Space2 is a large room close to CS classrooms that gives undergraduates a workspace with tables for group projects, and a few sofas. It is also equipped with three computers, a printer, a projector, and wireless Internet access. The department plans to add more resources once it assesses how the room is being used by students.

Professor Lee Osterweil's 320 course, "Introductions to Software Engineering," centers on the development of a large piece of software by teams over the course of the entire semester. The teams typically meet one or more times per week in order to coordinate, says Osterweil. One of the teams has found the new undergraduate space essential in getting their 320 project completed. "Company A" team meets at least twice each week in U-Space2. "We need a place to get 12 to 18 people together several nights each week," says Company A team. "Without the U-Space, where would our group meet?" Company A team consists of team manager Alex Jackson and teammates David Arbour, Mario Barrenechea, Tyler Bonci, Mark Chen, Jesse Clark, Louis DeMaria, Felipe de Mello, Josh Haan, Jake Luszcz, Tristan Peck, Nicolas Scarrci, and Ethan Wortzman.

"Company A" members work on their project in U-Space2.

Staff News

Graduate Program Manager **Leeanne Leclerc** received the Administrative Excellence Award from Dean Mullin of the Graduate School, in recognition of her superior performance and outstanding service to fellow employees, students, faculty, and visitors. This is one of four, campus-wide, in this inaugural year for this award.

Claire Christopherson received the Dean's Service Award recognizing her contributions to the College in her role as CS Business Manager. She was presented the award at the CNS Convocation in September.

Thijs deVries joined PRISMS as an Associate Software Engineer 2.

The Information Extraction and Systems Lab welcomed **Timothy Vieira** as an Associate Software Engineer 2.

Daniel Parker joined the Center for Intelligent Information Retrieval as a Senior Systems Manager.

ADJUNCT FACULTY – – – – – *continued from page 11*

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.

"We are extremely pleased to welcome these new adjunct faculty members," says Department Chair Andrew Barto. "Adjunct appointments enrich the department in many ways, creating new opportunities for both students and faculty to make connections beyond the department, the campus, and even the discipline of Computer Science."

RECENT PH.D. GRADUATES – – *continued from page 13*

derstand and utilize massive text collections when additional modalities are considered and modeled jointly with text.

Ting Yang: *Operating System Support for Modern Applications; (*Emery Berger and J. Eliot Moss, Advisors); May 2009; Software Engineer, Intel Corp.

Computer systems now run drastically different workloads than they did two decades ago. The enormous advances in both hardware and software technologies dramatically reshape applications and their behaviors, thus introducing more challenges to system resource management. However, existing general-purpose operating systems have not kept up. Their resource managers (CPU, memory and disk I/O) still work independently, leading to various performance issues. Garbage-collected applications suffer under memory pressure because the virtual memory manager does not provide enough information for them to adjust heap size. Interactive applications may lose responsiveness because one resource is overloaded.

To deliver robust performance, an operating system has to coordinate its resource managers, as well as cooperate with those in the user space (e.g. garbage collector and thread manager). To support garbage-collected applications, we present CRAMM, a system that enables them to adaptively adjust heap size to maintain high throughput by cooperating with the underlying virtual memory manager. To support highly interactive workloads, we present Redline, a system that uses lightweight specifications to drive CPU scheduling and to coordinate memory and disk I/O management to maintain system's responsiveness. Our experiences show that such coordination can dramatically improve application performance under heavy contention, sometimes by orders of magnitude.

Significant Bits

Newsletter of the
Department of Computer Science
College of Natural Sciences
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
Graduate Student Liaisons	Filip Jagodzinski, Borislava Simidchieva
Contributors	Ramesh Sitaraman, Bryant York, 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 June 2009 through October 2009. 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 ('85)
Dr. James P. Ahrens ('89)
Ms. Erica Asai ('00)
Drs. Anthony P. & Catherine J. Auger ('92)
Mr. Michal P. Cialowicz ('07)
Stephen Werntz & Kimberly Edgar ('86)
Andrew & Pamela Hannon ('01)
Mr. Hanumantha Kodavalla ('88)
Mr. Matthew D. Krajewski ('97)
Dr. David J. Lee ('81)
Karin & Andrew Merlino, Jr. ('86)
Ms. Terri A. Noyes ('83)
Mr. Charles N. Paliocha ('83)
Ms. Laura Anne Price ('87)
Louis & Mary Raboin ('78)
Mr. Lin Song ('01)
Mr. William H. Stevens III ('95)
Mr. Michael F. Sullivan ('85)
Keith & Amy Swan ('90)
Dr. Zhenlin Wang ('99)
Mr. Frederik Wijnen-Riems
Mr. Peter F. Willey

Corporate and Matching Gifts
AT&T
eBay CAN!
EMC
Google
IBM
LogicBlox
Microsoft Corporation
Raytheon
Yahoo!

Popplestone Fellowship
Anonymous–
In memory of Robin
Prof. Kristin Morrison

Paul Utgoff Memorial Fund
Prof. Andrew G. Barto
Marina & Philip Braswell
Ms. Cassandra L. Chew ('90)
Prof. Lori Clarke
Mr. Ding Gang ('00)
Prof. Kevin E. Fu
Dr. Robert A. Jacobs ('87)
Mrs. Ida N. Lauter
Prof. Victor Lesser
Prof. Sridhar Mahadevan
Mr. Michael J. Railey ('02)
Dr. Michael T. Rosenstein ('03)
Mrs. Alexandra Utgoff Taylor ('74)
Mr. Mathew S. Thoennes ('02)
Mrs. Karen Utgoff
Ms. Naomi Utgoff
Mr. Vadym W. Utgoff
Victor A. Utgoff
Mrs. Emily L. VanHassel