

# Significant **BITS**

Newsletter of the  
Department of Computer Science

## FACULTY

Welcome to  
Emery Berger,  
Oliver Brock, and  
Andrew McCallum  
page 3

## AWARDS

Jim Kurose receives  
NSM Service  
award  
page 5

## NEWS

Department  
celebrates 30th  
anniversary  
page 10

## CENTERS

CIIR tackles  
National Digital  
Library research  
page 9

## ALUMNI

Researcher Focus  
on Lisa Ballesteros  
alum matters page 7

## Challenges in designing autonomous, ubiquitous computing systems

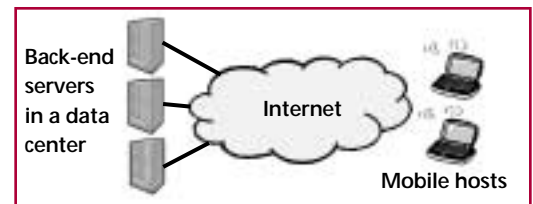
**O**VER THE NEXT FEW YEARS, NETWORKED MOBILE DEVICES WILL BECOME UBIQUITOUS: cellular phones with data access capabilities, Personal Digital Assistants (PDAs) with wireless networking interfaces, automobiles with onboard information systems, and others. Such devices will enable users to tap into a vast reservoir of data and computational resources and provide “anytime anywhere” access to information and computation. The Internet of the future (right) will consist of a large number of front-end mobile hosts that access storage and computational resources in back-end data centers. A distributed overlay network of edge servers acts as intermediaries between these front-end and back-end nodes. The distributed overlay network provides efficient, timely access to data and computation to front-end nodes.

Assistant Professor Prashant Shenoy’s research focuses on systems issues in realizing such an Internet architecture. Two specific research issues are discussed below: delivering highly dynamic web content to PDAs in a coherent fashion and the design of self-managing server clusters.

### Disseminating dynamic Web content to mobile hosts

Networked mobile devices such as PDAs have three characteristics that differentiate them from traditional wired hosts. First, these devices frequently move from one location to another. Second, they frequently get disconnected from the network due to either poor network coverage at a location or an explicit shutdown by the user. Third, mobile devices have limited resources and battery capacities, and techniques to enhance battery life by limiting power consumption on these devices are highly desirable.

Despite their limitations, mobile devices are used increasingly to access the Internet. For instance, they are used to access highly dynamic content such as stock quotes, real-time weather



Mobile hosts accessing resources from back-end servers over the Internet

and traffic information, and sports scores – content that changes frequently, sometimes as frequently as once every few seconds. Providing mobile devices with up-to-date versions of such data items requires frequent communication and good network connectivity. These requirements conflict with the intermittent connectivity and resource-poor nature of mobile hosts.

Professor Shenoy’s research group, the Laboratory for Advanced System Software (LASS), has developed several techniques to bridge this gap. “These techniques utilize resources at edge servers to provide timely and coherent versions of

*continued on page 4*

## Faculty awarded ITR grants

**T**HE DEPARTMENT RECEIVED THREE AWARDS in the National Science Foundation’s (NSF) third annual Information Technology Research (ITR) Program competition. Among this year’s recipients is Assistant Professor Prashant Shenoy who received an award for “System Support for Mobile Multimedia and Web Applications.”

*continued on page 2*

## FROM THE CHAIR

By Bruce Croft

**T**HE 2002-2003 ACADEMIC YEAR BEGINS with a number of new colleagues and a new facility in the Computer Science building. We are particularly excited by our three new faculty; Emery Berger from the University of Texas at Austin, Oliver Brock from Stanford (and a startup), and Andrew McCallum from CMU (and a startup). Their interests, in the general areas of high-performance systems software, robotics, and machine learning for language technologies, are described in more detail in articles in this issue. The broad range of possible intersections with current research in the Department creates many opportunities for new collaborative projects. This continues our run of successful hiring seasons and we are hoping for a repeat this year!

The new facility in our building is the common room, which was officially opened in the first week of classes. This space is a renovation of a room previously attached to the main office, and is designed to provide the computer science community with a place to eat, meet, and relax. The room is divided into an eating area and a 'comfy chair' area and has been very popular once people worked out where it was.

Another prominent new arrival is Chancellor Lombardi. His energetic presentations on the importance of research and

scholarship in the University, and on the role of the Amherst campus in the State, have created considerable enthusiasm and optimism in the Department. We expect to work closely with the Chancellor on a range of projects aimed at building on the basis of excellence we have established.



For those of you familiar with our part of the campus, the most significant change is the Engineering building being constructed to the South of the Computer Science building. Although it will block some views of the Holyoke Range, we look forward to the completion of the new auditorium that is part of the building.

Finally, and most importantly, we have two new classes of students joining the Department. Our graduate admission was the most selective ever, with over 1600 applications for 40 positions. A notable feature of this year's class is that 80 percent are U.S. citizens. On the undergraduate side, there were 850 applications for freshman admission, resulting in 60 new majors.

We are looking forward to a productive and stimulating year. The environment for doing groundbreaking multidisciplinary and collaborative research and teaching in areas such as bioinformatics, learning, and information analysis has never been better.

## ITR grants ..... continued from page 1

In another award, Associate Professor Shlomo Zilberstein (PI) and Professor Victor Lesser (CoPI) began a three-year ITR project this fall titled "A Formal Study of Coordination and Control of Collaborative Multi-agent Systems Using Decentralized MDPs." Professor Leon Osterweil (PI), and co-PIs Lori Clarke and Edward Riseman, received an award for their four-year project, "The Analytic Web."

The Department also received several prior ITR awards. Professors Donald Towsley and Jim Kurose and ECE Associate Professor Lixin Gao (Ph.D. '97) received an ITR grant for "Scalable Quality-of-Service Control for the Next Generation Internet: Fundamental Challenges and Effective Solutions." Associate Professors Eliot Moss, Chip Weems, and former Associate Professor Kathryn McKinley were also awarded a grant for "Dynamic Cooperative Performance Optimization." The following UMass Computer Science Alumni are leading or participating in NSF/ITR projects: Claire Cardie (Ph.D. '94), Edmund Durfee (Ph.D. '87), Claude Fennema (Ph.D. '91), Adele Howe (Ph.D. '93), Debra Richardson (Ph.D. '81), Daniel Rubenstein (Ph.D. '00), Tuomas Sandholm (Ph.D. '96), Bryant York (Ph.D. '81), Zhi-Li Zhang (Ph.D. '97), and Wei Zhao (Ph.D. '86).

Over the past three years of the ITR program, UMass has received 10 awards for approximately \$9.5 million, placing the University in the top 25 schools in the country. The ITR program was established to preserve America's position as the world leader of computer science and its applications. The NSF funds innovative, high payoff research, which explores new scientific, engineering, and educational areas in information technology.

## Who's that tutor?

Center for Knowledge Communication (CKC) graduate student Ivon Arroyo and Professor Carole Beal of Psychology have been invited to the November conference of the Massachusetts Computer Users in Education (MassCUE), the professional society for Massachusetts K-12 technology teachers. Their presentation will be on intelligent tutors AnimalWatch and Wayang Outpost, both collaborations with the Psychology Department and the School of Education. See page 8 for the full story.



## Department welcomes three new faculty

### Emery Berger

EMERY BERGER JOINED THE DEPARTMENT AS AN ASSISTANT PROFESSOR THIS FALL. "I'm excited to be here because there's a lot of synergy between myself and the existing systems group, especially with the work being done here in operating systems and compilers," says Professor Berger. "I've found the faculty here to be remarkably open and interested in collaboration, even across wildly disparate research areas. And I'm already enjoying working with some of the great graduate students in the Department."

Professor Berger will be starting the new Robust Software Systems Laboratory. "Modern software systems are brittle," says Berger. "Unanticipated workloads bring most applications to their knees. For instance, when many people try to visit a web site at the same time, the site often crashes instead of smoothly degrading its service." Research will focus on systems-based approaches to make robust software that performs well regardless of workload, architecture, or data access patterns. Berger is particularly interested in achieving robustness through greater cooperation between operating systems, runtime systems, and compilers.

Berger received his Ph.D. at the University of Texas at Austin. (his Ph.D. advisor was Kathryn McKinley, previously an Associate Professor of Computer Science at UMass). In his thesis work, he developed a number of high-performance runtime systems for modern programming languages, focusing on memory management. Berger notes that "current general-purpose memory managers do not

*continued on page 6*



### Oliver Brock

OLIVER BROCK, WHO RECENTLY JOINED THE FACULTY AS ASSISTANT PROFESSOR, has broad interests. "The computer science faculty at the University of Massachusetts Amherst has great strengths in many disciplines relevant to the field of robotics, such as artificial intelligence, computer vision, and machine learning," says Brock. "In addition, ongoing research efforts in robotics exhibit many synergies with my own work. Consequently, the Department is an ideal environment for the realization of my interdisciplinary research plans."

Brock's research focuses on algorithmic foundations for the perception, simulation, and interactions of virtual and physical worlds. Such efforts combine approaches from robotics, computer vision, computational biology, computer graphics, and dynamic and haptic simulation. "Progress at the intersection of these disciplines promises to have an impact on a diverse set of applications, ranging from simulated protein folding to virtual prototyping, and from medical applications such as remote surgery and surgical training to robotic assistants and tele-operated manufacturing," says Brock.

Brock's current research aims to enable complex robotic systems, such as humanoid robots, to act and interact autonomously and in complex and dynamic environments. Brock plans to continue his work on aspects of motion in the context of robotics. "To most people it seems like a trivial task to control a robot,

*continued on page 5*



### Andrew McCallum

ANDREW MCCALLUM, RECENTLY VICE PRESIDENT OF RESEARCH AND DEVELOPMENT AT WHIZBANG LABS, has joined the Department as a Research Associate Professor. He is active in research on machine learning and statistical methods applied to text. His particular research interests include Web mining, information extraction, document classification, finite state models, and learning from combinations of labeled and unlabeled data.

"After two and a half years in a research-based start-up, I believe more than ever in the ability of real-world problems to inspire great research - including significant algorithmic, modeling, and theoretical advances. Real problems are richly complicated, and, given the right approach, they can result in new fundamental work that might never have come to light without the inspiration of a tough problem," says McCallum. "I am returning to academia from industry because I love research, student advising, and teaching, and I haven't been able to get enough of those in industry. I also want the freedom to consider a wider array of problems and collaborations."

A driving goal of McCallum's research is building systems that can mine large bodies of unstructured text - such as the Web or collections of research literature - and find patterns, relations, and trends. Two of his previous projects did just that. One mined the Web for job openings; the other mined on-line document collections for computer science research papers and their citations. In both cases, the raw data was normalized, organized, and mined using machine learning methods. What resulted was automated analysis that could predict hiring trends by categorizing job openings in different states, or that could find the seminal papers on a research topic by analyzing the mined citation graph.

*continued on page 6*





## Challenges . (from page 1)

dynamic content to mobile end-hosts,” says Shenoy. “The key idea is to have edge servers (or proxies) synchronize content with the primary servers and then have the proxies push relevant content to the end-hosts (below).” As an example, consider a user who wishes to be notified every time the stock price of a certain company changes by \$1 or more, and also wants up-to-the-minute information about certain NFL games. The mobile host informs a nearby proxy of the user’s preferences; the proxy in turn is responsible for fetching up-to-date versions of this content from the corresponding servers and for pushing any update of interest to the mobile device. Observe that since a stock price may change more frequently than the score in a NFL game, the former needs to be refreshed more frequently than the latter. Shenoy’s group has developed adaptive techniques to dynamically determine the frequency of refresh for dynamic data items depending on their rate of change. Thus, if a stock becomes more volatile due to certain news events or if both teams start scoring more frequently in a NFL game, then the proxy dynamically detects an increase in the rate of change of these web pages and adjusts the frequency of refresh accordingly. The frequency at which dynamic content is refreshed from the server is chosen such that the user-specified error tolerances are not exceeded (changes of \$1 or more to the stock price is an example of the user-specified tolerance).

“These adaptive techniques involve a combination of proxy-pull and server-push for dis-

seminating updates to dynamic content from servers to proxies,” notes Shenoy. “Further, since the proxy only pushes those updates that are of interest to the mobile end-host, the communication on the client-proxy path is minimized.” Less frequent communication results in lower power consumption at the client and enhances battery life. If the mobile host gets disconnected, the proxy can intelligently detect this condition and use its storage resources to buffer all updates received during the disconnection period. Buffered updates are delivered to the end-client upon reconnection, which ensures that no interesting updates are missed even though the end-host may be temporarily disconnected. Finally, each proxy continuously tracks the geographical location of the end-host and intelligently hands over responsibility for serving the client to another proxy if the client moves to a new location. This ensures that a nearby proxy always services the end-client, resulting in lower network overhead and better response times. This proxy-based approach accounts for all three characteristics of mobile hosts: client mobility, intermittent connectivity, and client’s limited resources.

Shenoy’s group is implementing a prototype system, and developing several applications to demonstrate the effectiveness of their approach. One is a PDA-based stock ticker. The end-user specifies the stock items of interest and the desired error tolerance. The PDA registers these with the proxy, which then dynamically fetches updates to stock prices and delivers them

to the PDA. Disconnections and client mobility are handled as described above.

## Design of self-managing server clusters

A different set of considerations arises when designing back-end servers for Internet applications, which are often housed in data centers. These typically provide computational and storage resources for a large number of applications, each running on a cluster of servers. “Due to the sheer number of servers and storage systems in the data center, managing these systems becomes a key challenge,” says Shenoy. System administrators must deal with the difficult task of configuring large, complex systems so as to achieve high availability and performance. Data center management involves tasks such as

- configuration and provisioning of hardware resources for each application,
- continual tuning of the operating system software and provisioning of resources based on changes in the application workload, and
- handling failures of computing elements in the data center.

Companies are relying more and more on online applications. The cost of mis-configurations can be high since even a short down-time can result in substantial revenue losses. One possible approach to address this crucial problem is to make the data center autonomous or self-managing. A self-managing system automates the tasks of configuration, tuning, and failure handling. Human intervention is required only for unusual circumstances. This simplifies the administration of large servers, reduces administration costs, and most importantly, reduces the chances of human error.

Shenoy’s research group is investigating several techniques for designing self-managing data centers. These techniques in-

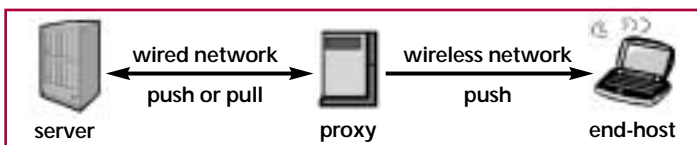
volve continuous online monitoring of the system to detect significant events such as a sudden workload change or the failure of a system component. These events in turn trigger system reconfigurations or failure handling mechanisms. “For instance,” says Shenoy, “if a newspaper web site sees a sudden increase in requests due to a breaking news story, a self-managing server can detect this workload change and automatically allocate additional server capacity or even additional servers to handle the increased workload.” Similarly, if some servers in the system become hot-spots while others are under-utilized, a self-managing system triggers load balancing techniques to rectify the imbalance. Shenoy’s research group combines ideas from control theory, machine learning, and non-linear optimizations to design such self-managing systems. Ongoing work focuses on

- dynamic resource provisioning based on workload changes,
- automated failure detection and recovery to make the system self-healing, and
- self-managing data placement relying on hot spot detection and online data migration.

His group is building a prototype self-managing server cluster based on the Linux operating system to serve as a test-bed for evaluating these techniques.

Professor Shenoy received his Ph.D. in Computer Science from the University of Texas at Austin in 1998, and that fall joined the Department. His Laboratory for Advanced Systems Software (LASS) conducts research in multimedia systems, operating systems, cluster computing, mobile computing, and content distribution networks. Professor Shenoy has received the NSF Career award (’00), an NSF ITR award (’02), an IBM Faculty Partnership Award (’00), and a Lilly teaching fellowship (’01).

Use of proxies to deliver dynamic content from servers to mobile hosts



## AWARDS

## Kurose receives NSM faculty award

**A**T THE UMASS COLLEGE OF NATURAL SCIENCE AND MATHEMATICS (NSM) fall convocation, Professor Jim Kurose received the outstanding faculty award for service. This award was established in 2000 to honor individual faculty members for their service contributions. Departments make nominations for these awards and the selections are made by a college-wide committee appointed by the NSM Dean. Kurose was nominated for the multi-faceted service contributions he has made to the field and to the University.

Professor Kurose was the founding Editor-in-Chief of the IEEE/ACM Transactions on Networking, widely regarded as the top journal in the networking field. Importantly, this journal is sponsored by both the IEEE and the ACM (the only such co-sponsored journal) – a unique collaboration that Kurose and the other co-founders worked hard to achieve. Professor Kurose has also been the Technical Program Co-chair of the three major conferences in his field: the IEEE INFOCOM Conference (1992),

the ACM SIGCOMM Conference (2000), and the ACM SIGMETRICS Conference (2000). He has served on a number of advisory or governing boards including the IEEE Communication Society Board of Governors (for two terms), the IEEE INFOCOM Standing Committee (since 1992), and the ACM SIGCOMM Technical Advisory Committee (for two terms). He has also served as the Director of Journals for the IEEE Communication Society.

In addition to his work within technical societies, Professor Kurose has also made outstanding service contributions at the University, the state, and national level. During and since his time as Chair of the Department, Professor Kurose has been a member of several state-level IT-related committees in Massachusetts. Nationally, he has served on a number of panels for the National Science Foundation and the National Academy of Sciences/National Research Council.



Jim Kurose

## Brock..... (from page 3)

because humans are so good at moving their own bodies. But if you try to explicitly specify the motion of each joint of your body as you climb up a flight of stairs, you will realize that this is not so trivial after all.” Methods for generating the motion of robots have to address a wide range of requirements, such as collision avoidance, balance, contact forces, and constraints imposed by the task. Some of the algorithms resulting from Brock’s research could find application in structural molecular biology. “Methods developed in the context of robotics have been shown to facilitate the understanding and modeling of motions and interactions of molecules,” says Brock, who intends to pursue this line of research here.

Before joining the Department, Brock was a Research Associate and Scientific Consultant at Stanford University’s Robotics Laboratory in the Computer Science Department. He worked with his former advisor, Oussama Khatib, on various aspects of elastic planning for robots with many degrees of freedom. He conducted research on motion generation, motion planning, and motion execution algorithms for humanoid robots, robot control, and computer vision. Prior to his most recent position at Stanford, Brock was a Lecturer and Research Scientist at Rice University. During that time, he also held the position of Founder and Chief Technology Officer of AllAdvantage.com, a pioneering Infomediary. An Infomediary – or information intermediary – is a trusted third party in the relationship between consumers and vendors.

Brock received his M.Sc. and Ph.D. in Computer Science from Stanford University, and his Diplom (B.S. / M.Sc.) was from the Technical University Berlin.

“I am looking forward to interactions and collaborations with students and colleagues, both in teaching classes and pursuing research projects,” says Brock.

## Who’s in the news?

The Department continues to attract attention from the national media. An article in the British science newsweekly *New Scientist* featured Professor **Paul Cohen’s** robot baby work. This work, in which a robot can learn the meaning of objects and words somewhat like a baby learns, was also profiled in an article in the *Edmonton Sun*. Cohen was interviewed by Todd Mundt of Michigan Public Radio for the *Todd Mundt Show*, a National Public Radio broadcast program. Cohen discussed artificial intelligence, “meaning,” and robots during his interview.

Airing on *ESPN’s SportsCenter*, Center for Intelligent Information Retrieval (CIIR) researcher **David Pinto** speculated on the probability that the L.A. Lakers would sweep the NBA final series against the New Jersey Nets. Touted as ESPN’s “resident expert,” Pinto’s interview was shot in the Department for the evening broadcast.

In the *Springfield Union-News*, **Paul Hudson** of the Center for Knowledge Communication (CKC) was noted in an article about Springfield fifth graders who visited the CKC lab where they viewed animation technology.

To learn about upcoming events,  
click “Department Calendar” at  
[www.cs.umass.edu](http://www.cs.umass.edu)

## Have a seat and enjoy the view

**T**HE ENTRANCE WAYS TO OUR BUILDING have become more colorful this season thanks to the efforts of the Department's volunteer gardeners. A small group of students, staff, and faculty have established a large garden bordering our front walk. This beautiful retreat is a welcome and relaxing diversion from day-to-day activities.

New gardeners and fresh ideas are always welcome. Please email Priscilla Coe at [coe@cs.umass.edu](mailto:coe@cs.umass.edu) to join the crew for events like our recent fall-bulb planting day.

The Department has been very generous in contributing to this effort, but your gifts will help purchase benches, plants, tools, and planters. We have provided an envelope for your contribution to go directly to the Campaign for Computer Science. Please specify "garden fund" on the envelope if you would like all or part of your gift to help enhance the garden.

There are many other current needs within the Department, and your support, in whatever form you choose, continues to make it an even better place. When you visit the Department, you will be able to see first hand where your contributions have gone and how they benefit everyone in Computer Science. Thank you!



### McCallum ..... (from page 3)

The same machine learning and pattern recognition techniques that work for these problems often also apply in other areas, such as bio-medical data modeling, network modeling, robotics, and optimization.

Before joining UMass, McCallum directed WhizBang's 30-person research lab in Pittsburgh, PA. He also held an adjunct faculty position at Carnegie Mellon University, where he co-advised several Ph.D. students and helped teach courses. Before WhizBang, he was a Research Scientist and Coordinator at Just Research (Justsystem Pittsburgh Research Center), where he spearheaded the development of methods for statistical text processing. There he also lead the project that created Cora, the domain-specific search engine over computer science research papers mentioned above. In 1996, he was a post-doctoral fellow at Carnegie Mellon University, where he worked with Sebastian Thrun on the Intelligent Building project and with Tom Mitchell on the WebKB project. McCallum graduated summa cum laude with a B.A. from Dartmouth College in 1989, and received his Ph.D. in computer science from University of Rochester in 1995, where he worked with Dana Ballard. His Ph.D. work, which was nominated for the ACM Best Thesis Award, did not focus on text, but rather

on models of short-term memory and selective perception for reinforcement learning agents and humans.

McCallum has over 35 research publications and several pending patents. He is on the editorial board of the Journal of Machine Learning Research, and has served on the program committees for many technical conferences, including IJCAI, AAI, ICML, NIPS, and UAI. He has given invited talks at MIT, Stanford, CMU, UT Austin, Brown, Xerox PARC, IBM Almaden, SRI, AT&T Research, and Google. In December, he will be giving a tutorial, "Information Extraction from the World Wide Web," at the Neural Information Processing Systems (NIPS) conference in Vancouver, Canada.

"I'm tremendously excited about joining the Department here at UMass," says McCallum. "I was involved with undergraduate education at Dartmouth, graduate student advising and research at CMU, and large project leadership at a startup - but each of these were separate, disjoint activities. Now at UMass I'm looking forward to creating the natural synergies between all three." McCallum continues, "With such world-class strength in machine learning, data mining, information retrieval, and theory, as well as such a strong tradition of multi-disciplinary collaboration, UMass is a great place for my work, for me to help form more bridges between these areas, and for me to contribute."

### Berger ..... (from page 3)

scale on multiprocessors, cause false sharing of heap objects, and systematically leak memory."

To address these problems, Berger developed the Hoard memory manager, which can improve application performance by up to a factor of 60. Hoard is widely used in both academia and industry, and is currently in use by AOL/Time-Warner, British Telecom, Novell, and Philips. He also developed a memory management system to support server applications that must tear down memory associated with terminated connections or transactions.

His system provides higher performance and lower memory consumption than current approaches.

His other research interests include multiprocessor scheduling, data locality, parallel and multithreaded programming, and the use of compiler technology for both optimization and error detection. In future research, Berger will continue to focus on solving problems with practical impact, applying theory when appropriate to guarantee system performance. As both software and hardware grow in complexity, Berger will focus on transparently improving application performance and developing tools that make the programmer's job easier.





# Alum Matters

A newsletter for alumni of the Department of Computer Science

## Breaking through the language barrier

**R**ECENT YEARS HAVE SEEN AN EXPLOSION in the amount of electronic and on-line information. Although much of the content is in English, the amount of non-English resources and the number of non-English speaking Internet users are growing steadily. Increased global awareness, multi-national collaboration, and interest in foreign media sources have driven a need for Cross-Language Information Retrieval (CLIR) systems that span language boundaries. These systems allow the submission of a query in one language (e.g. Spanish) and retrieval of documents in other languages (e.g. Arabic). The Center for Intelligent Information Retrieval (CIIR) was one of the first groups to explore this area and is still at the forefront of advances in CLIR research. Lisa Ballesteros (Ph.D. 2001), former student of Bruce Croft and now Clare Boothe Luce Assistant Professor of Computer Science at Mount Holyoke College, continues to investigate techniques that enable access to and management of multilingual media.

Ballesteros's research focuses on language independent approaches exploiting readily available lexical resources. She developed a cross-language retrieval system based on approximate translation via machine-readable dictionary, augmented with statistical techniques for reducing the effects of translation ambiguity. Despite a relatively unsophisticated translation resource, Professor Ballesteros was one of the first to

report a cross-language method that almost completely eliminated the effect of translation error from the retrieval process. "Dictionaries are more readily available than are other lexical resources, but there are still some language pairs for which no bilingual dictionaries exist," says Ballesteros. To address this issue, she became the first to publish results showing that transitive translation, which uses a third "pivot" language as an intermediate in the translation process, is a viable approach, producing acceptably accurate translation and subsequent retrieval. Her most recent work is investigating the use of statistical stemming for Arabic monolingual and cross-language retrieval. This is a language independent approach that measures the relatedness of stem-class pairs via their co-occurrence in fixed text windows.

"CLIR tools can facilitate routine tasks of groups such as multi-nationals, information services, and government organizations (for instance, the Patent Office). These organizations routinely hire banks of multilingual speakers and translators to

cull through vast amounts of foreign-language text to locate important information," says Ballesteros. "Furthermore, the events of September 11 have increased interest on the part of security organizations that are looking for ways to automatically identify items of interest in foreign language collections." Demand is also increasing for CLIR techniques to manipulate multimedia data

including voice, images, and video. The growth of mobile accessory use has also led to challenging research questions regarding the development of mobile personal information retrieval systems. Such a system would exploit GPS information and other multimedia data to fulfill a specific information need given the user's current location. For example, a tourist may need to find a drug store that is nearest her current location. Unlike traditional IR systems in which results are generated using a global approach, mobile IR systems will need to focus on local context.

"The information technologies of the future will need more than the ability to cull through massive amounts of information," says Ballesteros. "They must also offer a wide range of tools for information analysis and manipulation; tools that facilitate discovery and organization, and that support the many new ways in which people interact with information."

Professor Ballesteros is spending her sabbatical at the University of Sheffield and at UMass, Amherst. She will be working on two main projects in the coming year. The first is developing corpus-based metrics that can be used to guide parameter selection for statistical stemming of different languages and corpora. The second is to develop a technique that relies on statistical analysis of contextual cues to infer translations for out-of-vocabulary (OOV) words. OOV words generally include special vocabulary such as technical terms that typically do not occur in bilingual dictionaries. An OOV solution would be useful for a wide range of tasks, including lexical acquisition and mobile personal IR, where a person may be unfamiliar with the correct foreign language term or expression to use in a specific context.



Ballesteros

## Intelligent tutors advance

**T**HE CENTER FOR KNOWLEDGE COMMUNICATION (CKC) has recently received two awards to continue research and development of innovative approaches to computer-based learning. One is embodied in a system called Wayang Outpost and is designed to help redress the shortage of women entering math and science careers. The other, called Rashi, integrates inquiry-based approaches to science teaching pioneered at Hampshire College into an advanced instructional system.

Wayang Outpost follows the success of CKC's AnimalWatch, an intelligent tutor that teaches mathematics to grade schoolers. Sponsored by the National Science Foundation's Program for Women and Girls, AnimalWatch puts the student in the role of a wildlife biologist studying endangered species, an approach that encourages girls to contextualize the use of math in an exciting scientific endeavor (see <http://sophia.cs.umass.edu/AWE/>). Wayang Outpost, funded by a grant from NSF, has a similar goal but is aimed at a higher age level – high school girls preparing for the SAT's. Carole Beal, Department of Psychology, is the principal investigator of both projects and Research Associate Professor Beverly Park Woolf is a co-PI.

High-stakes achievement tests such as the SAT, GRE, and MCAS have become increasingly important. A student's performance can have a significant impact on access to future educational opportunities. Concern is also growing that such tests simply exacerbate existing group differences, and put female students and those from traditionally underrepresented minority groups at a disadvantage. New approaches are required to help all students perform their best on these tests. The Wayang tutoring system has gender in mind while teaching for high-stakes math achievement tests. It identifies methods and teaching strategies (e.g., visual or analytic) preferred by each gender and then customizes the system responses to the individual student. Set in Borneo and studying the orangutan, Wayang already provides web-based instruction in a Macromedia Flash animated classroom. Mathematics SAT problems are presented in Flash movies featuring an animated character based on the traditional Indonesian Wayang art form of shadow puppetry. When the student answers incorrectly, or requests help, the puppet provides step-by-step guidance in the form of Flash animations with audio. A server-side engine keeps track of students and decides what problems and help to provide.

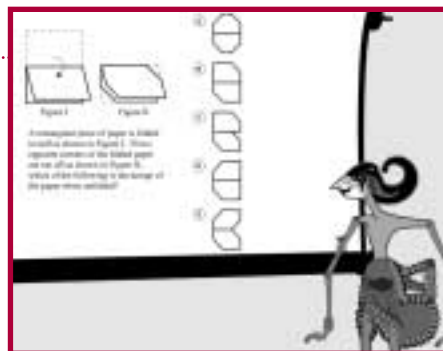
Why an intelligent tutor? One of the advantages is its ability to represent the knowledge of individual users, and to select problems and help to match the user's current level of understanding. This student model is continually updated as the student works and is used to adapt instruction to the user. Wayang also models information about the student in areas that have been shown to predict performance on math achievement tests, specifically, spatial ability and fact retrieval speed. Because gender and other group differences have been observed in these component skills – with implications for math test performance – Wayang Outpost will include on-line assessments of these cognitive skills as well as activities designed to enhance these abilities.

Wayang will make intelligent decisions about problem selection, help selection, and probably emotions selection for the helping character. Research has shown that important student characteristics, such as math ability, gender, spatial ability and fact retrieval, are relevant when students solve geometry problems.

The system will dynamically adjust problem features while working with a student. These include visual-estimation or analytic help (two strategies that past research suggest male and female students use differently), the level of detail in the explanations, problem difficulty, and even the emotion to display in the character providing help. Machine learning classifiers will help the system decide if the amount and the quality of the pedagogical content presented will be effective depending on student and content characteristics. Wayang will be evaluated in the fall with random decisions about help and problem selection. The objective is to collect data that will be used to learn these pedagogical decisions based on actual student interactions with the system.

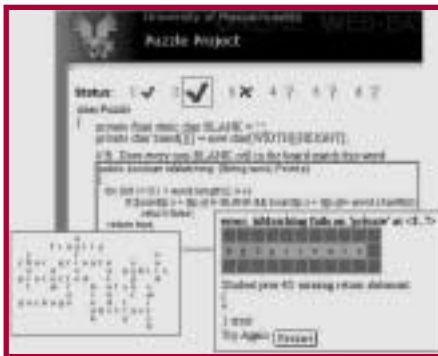
The second project continues a collaboration with Hampshire College researchers exploring the integration of inquiry-based science methods, used systematically in the Hampshire curriculum, into advanced instructional applications. Woolf is principal investigator. During the first phase, CKC researchers built a prototype system called Dr. Geo that simulated a geology field trip. Students can enter observations made from interactive pictures and diagrams, generate hypotheses to explain observed phenomena, and gather evidence from the simulation to support or refute their hypotheses.

Under a new grant from the U.S. Department of Education's Fund for the Improvement of Post-Secondary Education, CKC researchers are generalizing the techniques developed for Dr. Geo to work across a variety of domains including biology and civil engineering. Students will use a suite of interactive tools to record their



Wayang Outpost's interactive characters such as traditional Indonesian Wayang shadow puppets (above) help the student solve difficult spatial reasoning problem such as those found on the Scholastic Aptitude Test (SAT). (Character of shadow puppet developed by Ari Bergman)

observations, gather scientific data, and graph their hypotheses. One of these tools is a derivative of the graph-based Belvedere system for scientific reasoning developed by Dan Suthers (Ph.D. '93). Called Rashi, after an 11th century Talmudic scholar, it will be field tested by co-designers at Hampshire, the University of Rhode Island, and the UMass Geology and Civil Engineering Departments. This project dovetails with several NSF-sponsored projects at Hampshire exploring the impact of inquiry-based pedagogy.



CS Homework and possible system responses.



## CIIR tackles National Digital Library research

**T**HE CENTER FOR INTELLIGENT INFORMATION RETRIEVAL (CIIR) has been awarded funding for two projects working on the National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program sponsored by the National Science Foundation.

Assistant Professor James Allan has been working as principal investigator on the core integration effort of the NSDL for over a year. That effort is a collaboration among several organizations, among them: Columbia University, which is developing support for rights management and access control; Cornell University, focusing on the core storage for the NSDL's global metadata; and University Corporation for Atmospheric Research (UCAR), which is providing overall management of the core integration project. The CIIR's role is to provide search and discovery services for the on-line library.

The CIIR has been focusing on providing fundamental search capability for the library, in a way that mixes metadata search with high-quality content-based search. Much of the work to date has focused on developing protocols to communicate between the various partners – e.g., so that the CIIR can access the contents of the metadata repository at Cornell in order to find out what items exist in the library. The process has also included the definition and implementation of a new NSDL query language for use by NSDL components (not intended for human searchers to use directly).

Future work on search and discovery will include a more robust implementation of some aspects of indexing and search, in order to scale to substantially more metadata records and simultaneous searchers (the NSDL project plans to support millions). The discovery component of the NSDL will be developed over the next few years, including capabilities such as automatic addition or correction of metadata, location of related material across different subcollections of the library, and concept browsing for more elaborate search capabilities.

In collaboration with Research Associate Professor Beverly Woolf and Professor Victor Lesser, the CIIR is also investigating methods for discovering and then using the topic structure of course material that is available on-line. The results of this work might be used to build courses automatically, to recognize when critical material is absent from the NSDL, or to help a teacher construct a course out of existing material.

In another NSDL project that was just awarded, Professor Bruce Croft is partnering with the Information Institute of Syracuse University (IIS) and the Wondir Foundation to enhance the NSDL by merging the information retrieval (IR) and digital reference components. By combining these functions, users can find answers to their questions regardless of whether those answers come from documents in NSDL collections or from experts accessible through the NSDL's virtual reference desk.

The combination of IR and digital reference provides powerful possibilities, with each field complementing the other. In

## CCBIT news

**I**N NEWS FROM THE CENTER FOR COMPUTER-BASED INSTRUCTIONAL TECHNOLOGY (CCBIT), CKC's sister lab, the Online Web-based Learning (OWL) system continues to expand. Last year, Professors Robert Moll, Allen Hanson, and Wendy Lehnert joined with CCBIT on an NSF start-up award to integrate OWL into the introductory Java programming class, CS121. Christopher Eliot (Ph.D. '96), one of the course instructors, designed a number of programming exercises that allowed students to submit code and receive immediate constructive feedback. A follow-up proposal to expand this effort will be funded by a two-year NSF award starting in January. Meanwhile the OWL system was successfully applied in chemistry by Thomson Publishing. Over 5,000 students from across the nation used OWL regularly last year, and 25,000 students are expected to use the system this year. The web-based homework system carries the UMass logo and produces a per-user royalty for the University.

In other news, CKC Director and Research Associate Professor Beverly Park Woolf was nominated for a 2002 World Technology Network (WTN) award for education. The WTN explores emerging technologies that will have the greatest impact upon industry and society in the first years of the 21st Century. These range from information and communications technology to new materials, new energy and new space technologies, and biotechnology. Dr. Woolf attended a two-day gathering of about two hundred innovative technology leaders, including such notable speakers as Sir Arthur Clark, a Nobel Peace Prize nominee.

terms of IR, experts and human intermediation developed in digital reference can handle questions currently beyond the scope of the IR system, or even the collection being searched. The technical focus is to develop techniques for *query triage*, a process where queries will be routed automatically to the appropriate IR or human resource. This research will involve the application of machine learning and language modeling techniques to large query logs collected in an operational environment.

The CIIR's partners bring broad experience to the project. The Information Institute of Syracuse (IIS) is the umbrella organization for a number of highly visible and widely successful digital education information services. The IIS administers and sponsors the ERIC Clearinghouse on Information & Technology and related projects for educators at all levels, including the award-winning, Internet-based 'AskERIC' question and answer service. The Wondir Foundation's mission is to provide a free, universal, integrated, independent, and unbiased search service, combining search and community in ways possible only in a nonprofit organization. Wondir will combine Search and Community – two pillars of the Internet that have not yet lived up to their potential – by offering the instant help of the millions of experts and enthusiasts online at any one time, along with a broad metasearch of web sources and valuable 'invisible' web resources.

## Department celebrates 30th anniversary

ON OCTOBER 9, THE DEPARTMENT CELEBRATED THE 30TH ANNIVERSARY OF THE PH.D. PROGRAM with a research open house. Chancellor John V. Lombardi welcomed the group with his opening remarks. A poster and demonstration session showcased the Department's research areas, and faculty, alumni, and industrial partners attended a celebratory lunch.

As part of the day's events, CS alumni on a research panel reflected on a career in computer science. Alumni included Yuan-Chieh (Randy) Chow (Ph.D. '77) of the National Science Foundation and the University of Florida, Laura Dillon (Ph.D. '84) of Michigan State University, A. Anthony Gee (B.S. '90) of Carthage Venture Partners, and Wei Zhao (Ph.D. '86) of Texas A&M University.

The new Microsoft Education Laboratory was also unveiled during the research open house. Scott Williams, Microsoft Research Program Manager of University Relations, and Assistant Professor Prashant Shenoy spoke to the group about the new lab and Microsoft partnership. As a result of Shenoy's proposal submission for the lab equipment, Microsoft donated 35 machines and a server for the Windows-based Education Lab.

To continue the celebration, research labs such as the Center for Intelligent Information Retrieval (CIIR) and the Laboratory for Advanced Software Engineering Research (LASER) held individual receptions for their alumni at the end of the day. Information and photos from the anniversary celebration can be found at: <http://www.cs.umass.edu/openhouse>.



**Alumni:** If you didn't receive notice of this event via a mailing or an email, then we don't have your current contact information. Please send us updates to [alumni@cs.umass.edu](mailto:alumni@cs.umass.edu), so we can invite you to future events.

## UMass hosts language modeling workshop

THE UMASS CENTER FOR INTELLIGENT INFORMATION RETRIEVAL (CIIR) hosted the Workshop on Challenges in Information Retrieval and Language Modeling held in the Department this fall. Language modeling is a new approach to information retrieval that was introduced to the field by Jay Ponte (Ph.D. '98) and Professor Bruce Croft in 1998. That orig-

inal work has sparked a large number of papers from a wide range of research groups throughout the world.

This workshop brought together a group of leaders in information retrieval, language modeling, and related fields. They spent two days talking about the major open challenges in the field. An important component of the work-

shop was to talk about how statistical language modeling techniques might address the challenges. The attendees are working on a final report that should be available by the end of the fall.

UMass and Carnegie Mellon University organized the event. UMass organizers included Professor Bruce Croft, Assistant Professor James Al-

lan, and Research Assistant Professor R. Manmatha. CMU organizers included Associate Professors Jamie Callan (Ph.D. '93) and John Lafferty. UMass and CMU are partners on a language modeling research project sponsored by the Advanced Research and Development Activity (ARDA), the organization that also sponsored the Workshop.

## Faculty News

**Rod Grupen**, Director of the Laboratory for Perceptual Robotics, was promoted to full Professor from Associate Professor. ■ Professor **Jim Kurose** was recently honored with the IEEE Communication Society Service Appreciation Award. In addition, he gave a keynote address at the IEEE International Conference on Computer Network and Mobile Communications and also completed a National Academy of Science Panel on Networking Research Horizons. ■ Distinguished Professor **Arnold Rosenberg** was appointed to the Editorial Board of *IEEE Transactions on Computers*, the flagship journal of the IEEE Computer Society. He has also been appointed to the Steering Committee of the 2003 IEEE Workshop on Advances in Parallel and Distributed Computation Models. ■ During the summer, Professor **Paul Cohen** was a fellow of the EPSRC (Engineering and Physical Sciences Research Council, UK) visiting the Math Department of Imperial College, London. He also gave a tutorial on Empirical Methods for Agent Systems at the Autonomous Agents and Multi-Agent Systems (AAMAS) conference in Bologna, Italy in July. ■ The paper "Cooperative Leases: Scalable Consistency Maintenance for Content Distribution Networks," was designated as an award paper at the 2002 World Wide Web conference. The paper was co-authored by graduate students **Anoop Ninan** and **Puru Kulkarni**, Assistant Professor **Prashant Shenoy**, and Professor **Krithi Ramamritham**. The paper has also been invited to appear in a special issue of the *IEEE Transactions on Knowledge & Data Engineering (TKDE)* journal. ■ Professor **Victor Lesser** gave the plenary speech, "Evolution of the GPGP/TAEMS Domain-Independent Coordination Framework" at the First

International Joint Conference On Autonomous Agents & Multi-Agent Systems (AAMAS 2002) in Bologna, Italy. ■ Associate Professor Shlomo Zilberstein was elected conference co-chair for the 14th International Conference on Automated Planning & Scheduling (ICAPS) to be held in Vancouver, Canada in 2004. ICAPS is the premier international conference in the area of Automated Planning and Scheduling, and one of the leading conferences in AI. ■ Associate Professor **Paul Utgoff** gave the keynote talk at the Workshop on Development of Representations, held in conjunction with the Nineteenth International Conference on Machine Learning, in Sydney Australia. ■ Associate Professor **Ramesh Sitaraman** was the keynote speaker at the Packet Video conference, and an invited speaker at Carnegie Mellon University, Boston University, and MIT.

## Visitor News

**Lisa Ballesteros** (Ph.D. 2001), from Mount Holyoke College, is a Visiting Assistant Professor with the Center for Intelligent Information Retrieval (CIIR). ■ **Patricia Riddle** and **Michael Barley**, both Faculty at the University of Auckland in New Zealand, are Visiting Faculty with the Knowledge Discovery Laboratory (KDL). ■ Amherst College Assistant Professor **Scott Kaplan** is spending his sabbatical as a Visiting Researcher working with Associate Professor **Eliot Moss**.

## Research News

Senior Research Fellow **Chris Hoffman** recently joined the Object Systems Laboratory. ■ The Laboratory for Software Engineering Research (LASER) welcomed Senior Research Fellows **Rich Sutton** and **Barbara Lerner**. ■ **Daniel Corkill** joined the Multi-Agent Systems Laboratory (MAS) as a Senior Re-

## Nico Spinelli retires

After nearly three decades at UMass, Professor D. Nico Spinelli took advantage of the early retirement program and retired this spring. Spinelli joined the Department as a tenured Professor in 1973. His research centered on the organization and development of the mammalian visual system. In combining experimental and theoretical approaches to understanding brain function, he was a pioneer in what is now widely known as Computational Neuroscience. Spinelli's 1970 model showing how the nervous system could implement a content addressable memory system was an early contribution to a collection of influential neural network models of memory. His innovative experiments examined, among other things, the influence of early visual experience on the developing visual system.

Spinelli was an active advisor in the Bachelor's Degree with Individual Concentration (BDIC) program at UMass. BDIC is a unique program that offers highly motivated and self-directed students the opportunity to design their own major. Prior to joining UMass, Spinelli was an Assistant Professor of Psychiatry at Stanford University. He received his M.D. from the University of Milan, Italy in 1958.



Spinelli

search Fellow. ■ The Knowledge Discovery Laboratory (KDL) welcomed **Amy McGovern** (Ph.D. 2002) as Senior Postdoctoral Research Associate.

## Student News

**Victor Lavrenko**, CIIR graduate student, won best student paper at the 24th European Colloquium on IR Research (ECIR '02). His paper, "Optimal Mixture Models in IR," was presented at the Glasgow, Scotland conference this spring. ■ **David Gucwa**, a Computer Science undergraduate student, won first place in the 17th Annual Mathematics Competition run by the UMass Department of Mathematics and Statistics. CS undergrad, **Ning Zhang** earned honorable mention in the competition. ■ Undergraduate student **Debangsu Sengupta** was awarded an Honors Research Fellowship

by the Commonwealth College. Sengupta is working with Research Assistant **Professor R. Manmatha** on his honors project.

## Staff News

**Paul Oh** joined the Center for Computer-Based Instructional Technology (CCBIT) as the Humanities and K-12 Project Coordinator. ■ The Knowledge Discovery Laboratory (KDL) welcomed **Agustin Schapira** as a Senior Software Engineer. ■ **Leslie Marsland** received her Associates Degree in Science in Business Administration from Springfield Technical Community College (STCC) this spring. Marsland, the Administrative Assistant for LASER, is also a Notary Public. ■ **Tyler Trafford** recently joined the Networks Laboratory as an Associate Staff Programmer.





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

## Significant Bits

NEWSLETTER of the  
DEPARTMENT OF COMPUTER SCIENCE  
at the UNIVERSITY OF MASSACHUSETTS, AMHERST

140 Governors Drive  
University of Massachusetts  
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](http://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](mailto:bits@cs.umass.edu).

Art Director ..... Karen Chrisman, North Haven Design  
Faculty Liaisons ..... Rick Adrion, Bruce Croft,  
David Jensen, Prashant Shenoy  
Staff Liaison ..... Jean Joyce  
Graduate Student Liaison ..... Hannah Blau  
Contributors ..... Lisa Ballesteros, Terri Gauthier,  
David Hart, Thom Kendall,  
Laura Macsuga

### GIFTS

## Thanks for your support

**T**HE FOLLOWING ALUMNI AND FRIENDS have actively supported the Department of Computer Science from April 2002 through July 2002. Such financial support is greatly appreciated and helps maintain a world-class instructional and research program. Contributions of alumni and friends help to fund important special activities that are not supported through the state budget.

Gary & Cheryl Almeida  
Frederick & Dale Anderson  
Dr. Krishnamoorthy Arvind '91  
Dr. Kevin D. Ashley '88  
Mr. John V. Bellissimo  
Charles & Merrill Belmer  
Mr. Rahul Bose '89  
Mrs. Anne F. Burgess  
Ms. Valerie I. Congdon '77  
Ms. Aiqun Du '98  
Peter & Vesna Flanagan

Mr. Craig L. Fournier '91  
Thomas & Susan Fulchino  
Frank & Patricia Fusaro  
Mr. Steve J. Hoogheem '79  
Don and Nancy Irwin '82  
Stephen & Karyn Labonte  
Eric & Amy E. Ludlam '93  
Mr. Andrew E. Merlino, Jr. '86  
Mr. Mark S. Michalski '81  
Mr. Paul D. Moineau '95  
John & Laura Nalesnik

Those interested in helping the Department should send a check made out to the University of Massachusetts to:  
Department of Computer Science  
140 Governors Drive  
University of Massachusetts  
Amherst, MA 01003-9264

Please state that your gift is *restricted to Computer Science*.

Kenneth & Sybil O'Connell  
Mr. Robert A. Piazza '86  
Paolo M. Piselli '99  
Victor & Kim Quinn  
Wilson & Elizabeth Rawlins  
Ms. Stephanie M. Rice '94  
Mr. Michael E. Rose '85  
Arnold & Susan Rosenberg  
Seth & Rebecca Rosenberg '77  
Ms. Kathryn Schlosky '82  
Olga Shvartsman

Dr. Elliot M. Soloway '78  
Mr. Dan Upp  
Mrs. Alexandra Utgoff Taylor '74  
Courtney Wade  
Mrs. Nancy W. Wasiuk '79  
Mr. David S. White '93

**Matching Gifts:**  
Hewlett-Packard Company  
Lockheed Martin  
Wellington Management Co.



The UMass Alumni Association provides a connection between alumni, faculty, friends, and the University. Stay connected by becoming a Member.  
**Membership matters.**

800-456-UMASS  
(8627)  
[umassalumni.com](http://umassalumni.com)