

THE LINK

News from the School of Computer Science
Issue 4.0 / Spring 2009

Carnegie Mellon

Head of the Class

RESEARCHERS AT CARNEGIE MELLON
AND PITT ARE UNTANGLING THE
BLACKBOARD JUNGLE

Also Inside:

POKERBOTS ADVANCE AI
SURROUNDED BY DATA
EVOLVING HORNOSTEL'S VISION



News from the School of Computer Science

The Link provides a mosaic of the School of Computer Science: presenting issues, analyzing problems, offering occasional answers, giving exposure to faculty, students, researchers, staff and interdisciplinary partners. The Link strives to encourage better understanding of, and involvement in, the computer science community.

Editor-in-Chief

Randal E. Bryant

Managing Editor/Writer

Jason Togyer

Contributing Writers

Tina Carr, Karen Hoffmann, Meghan Holohan, Philip L. Lehman, Byron Spice

Photography

Ken Andreyo, Maher Attar, David Bachman, Glenn Brookes, Wade H. Massie, Sanjiv Singh, Andrea L. Zrimsek

Graphic Design

Melissa Stoebe
Communications Design Group

Office of the Dean

4305 Newell-Simon Hall
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213

Randal E. Bryant, dean

Tina Carr (HNZ'02), director of alumni relations

Philip L. Lehman (CS'78, '84), associate dean for strategic initiatives

Byron Spice, codirector of public relations

Jason Togyer (HS'96), writer/editor

Anne Watzman, director of public relations

Carnegie Mellon University

412-268-7884
TheLink@cs.cmu.edu

Carnegie Mellon University does not discriminate and Carnegie Mellon University is required not to discriminate in admission, employment, or administration of its programs or activities on the basis of race, color, national origin, sex or handicap in violation of Title VI of the Civil Rights Act of 1964, Title IX of the Educational Amendments of 1972 and Section 504 of the Rehabilitation Act of 1973 or other federal, state, or local laws or executive orders.

In addition, Carnegie Mellon University does not discriminate in admission, employment or administration of its programs on the basis of religion, creed, ancestry, belief, age, veteran status, sexual orientation or gender identity. Carnegie Mellon does not discriminate in violation of federal, state, or local laws or executive orders. However, in the judgment of the Carnegie Mellon Human Relations Commission, the Presidential Executive Order directing the Department of Defense to follow a policy of, "Don't ask, don't tell, don't pursue," excludes openly gay, lesbian and bisexual students from receiving ROTC scholarships or serving in the military. Nevertheless, all ROTC classes at Carnegie Mellon University are available to all students.

Inquiries concerning application of these statements should be directed to the Provost, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, telephone 412-268-6684 or the Vice President for Campus Affairs, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, telephone 412-268-2057.

Carnegie Mellon University publishes an annual campus security report describing the university's security, alcohol and drug, and sexual assault policies and containing statistics about the number and type of crimes committed on the campus during the preceding three years. You can obtain a copy by contacting the Carnegie Mellon Police Department at 412-268-2323. The security report is available through the World Wide Web at www.cmu.edu/police/.

Obtain general information about Carnegie Mellon University by calling 412-268-2000.

Produced for the School of Computer Science by the Communications Design Group, March 2009, 09-412.

©2009 Carnegie Mellon University, all rights reserved. No part of this publication may be reproduced in any form without written permission from the Office of the Dean, School of Computer Science.

CALENDAR OF EVENTS

All events to be held at the Carnegie Mellon University campus in Pittsburgh, unless otherwise noted.

April 5

Sigbovik 2009: The Associate for Computational Heresy Special Interest Group

Newell-Simon Hall

April 16–18

Spring Carnival

April 17

15th Annual MoBOT Races

Wean Hall, front entrance

12 noon

April 18

SCS and ECE Spring Carnival Alumni Reception

Perlis Atrium, Newell-Simon Hall

1 to 2:30 p.m.

April 23

A Celebration of Teaching

Rangos Hall, University Center

May 6

Hank (Suz Chi) Wan Memorial Lecture

Marc Abrahams, founder, Ig Nobel Awards

Location TBA

May 16–17

Commencement Weekend

June 16–17

2nd Annual Game Summit

Entertainment Technology Center,
700 Technology Drive

October 29–31

Homecoming Weekend

FEEDBACK LOOP

You've got questions. And answers, too. We want to hear both. If you've got a comment about something you've read in The Link, or if you'd like to make a suggestion or submit a news item, please call us at (412) 268-8721, email TheLink@cs.cmu.edu, or write to The Link Magazine, Office of the Dean, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. Constructive criticism is thoughtfully considered. Rude retorts are folded, spindled and mutilated.

—Jason Togyer (HS'96), managing editor



10 / Head of the Class

The Pittsburgh Science of Learning Center—a joint project of Carnegie Mellon and the University of Pittsburgh—is building the world’s largest repository of educational data. Mining that information is providing new insight into the ways that children and adults learn.

By Jason Togyer



15 / Winning Hand

Pokerbots like the ones being developed by Tuomas Sandholm are a likely bet to revolutionize artificial intelligence, just as chess-playing computers did two generations ago.

By Byron Spice

Issue 4.0 / Spring 2009

2 / From the Dean

3 / Creating the Gates and Hillman Centers

Some people call it revolutionary. But architect Mack Scogin says the design of the new Gates and Hillman Centers is really an evolution of Henry Hornbostel’s vision.

5 / On Campus

SCS goes down on the farm, Scarab goes to Hawaii, the McWilliams Center reaches for the stars and Qatar’s Education City hires a “roboceptionist.”

9 / In the Loop

Bruce Nelson Professor of Computer Science Manuel Blum is looking forward to the next big developments in technology—and a fresh pot of coffee.

18 / Research Notebook

SCS Dean Randy Bryant kicks off a new Link feature with an overview of data-intensive supercomputing.

21 / Alumni Director’s Message

22 / Alumni Snapshots

23 / SCS in the News

Cover Image:

Ken Koedinger of the Human-Computer Interaction Institute is co-director of the Pittsburgh Science of Learning Center, a partnership between Carnegie Mellon and the University of Pittsburgh that was funded five years ago with the help of a \$25 million grant from the National Science Foundation. PSLC researchers are using advanced data-collection and data-mining techniques to understand how students can develop deeper strategies for learning languages, math and science. That’s Carnegie Mellon’s Collaborative Innovation Center and Pitt’s iconic Cathedral of Learning in the background. (Ken Andreyo photo)



Building, Adapting and Enduring

A large delegation from Carnegie Mellon traveled to Doha, Qatar, recently for the dedication of the new building for the Carnegie Mellon Qatar campus. The building can only be described as spectacular, larger than any building on the Pittsburgh campus, with an amazing sense of openness and space.

I was pleased to see the dedication of our faculty, staff and students to recreating in Doha our unique form of computer science education, combining deep thinking with real-world problem solving, in a part of the world with an educational system that's been more traditionally based on rote memorization. Both the building and the social and intellectual environment have proved so compelling that many of our recent graduates return to help students with their assignments.

Meanwhile, back in Pittsburgh, construction of the Gates Center for Computer Science and the Hillman Center for Future-Generation Technology continues, and both buildings move toward completion by the end of this summer. These buildings also will be spectacular, but in ways that are incomparable to the Qatar building. The Pittsburgh buildings are strikingly bold and innovative in their design, yet integrate well into the rest of campus.

The role of the buildings in forming a connection to the rest of campus, including the Randy Pausch Memorial Bridge between Gates and the Purnell Center, mirrors the way that computer science and computer technology play pivotal roles in connecting the intellectual endeavors of all of Carnegie Mellon.

Randy used to say that if there's an elephant in the room, it's important to talk about it. In that spirit, let me talk about how the current worldwide financial crisis is affecting the School of Computer Science and Carnegie Mellon University.

We are not immune to the hardships everyone is facing. The university's endowment is down, many of our industry sponsors have been forced to cut back their support, students and their families are finding it difficult to pay the high cost

of education and funding from the U.S. government has been at best flat over the past several years.

On the other hand, we have reason to believe that we will continue to thrive as an organization. On the whole, Carnegie Mellon depended on its endowment income for just 6 percent of its operating budget—a relatively small amount. Our research has been funded mostly by federal grants, and there are signs that this source of income will become stronger, in part due to some of the stimulus initiatives.

Our application rate for undergraduate, master's and Ph.D. programs remains very high. Students appreciate the value that a Carnegie Mellon computer science education can provide. We are, of course, evaluating many programs and looking for ways to reduce our expenses, but overall we feel confident we will continue to flourish.

The people in our organization have demonstrated a remarkable ability to seek out and adapt to new opportunities. This has allowed us to evolve with the times, without compromising our ability or desire to conduct world-class research and education.

Randy Bryant

Carnegie Mellon in Qatar celebrated the official opening of its new building in Education City on Feb. 22. H.H. Sheikha Mozah Bint Nasser Al-Missned, chair of the Qatar Foundation, cut the ceremonial ribbon. She was joined on stage by the entire Carnegie Mellon Qatar student body, along with Carnegie Mellon Qatar Dean Charles E. Thorpe, Carnegie Mellon President Jared L. Cohon and Qatar Foundation President Mohammed Fathy Saoud.



MAHER ATTAR PHOTO



Scogin: Buildings' modern design reflects the Hornbostel heritage of innovation



JASON TOGYER PHOTO

Looking to the Future, Respecting the Past

Genius—Thomas Edison said—is one percent inspiration and 99 percent perspiration.

Architecture—Mack Scogin says—also relies more on old-fashioned hard work than on bolts of creativity that descend like lightning from the heavens.

“You first have to establish all of the knowns—all of the necessities of the project,” says Scogin, co-principal of Atlanta-based Mack Scogin Merrill Elam Architects, designers of the Gates Center for Computer Science and the adjoining Hillman Center for Future-Generation Technologies, slated to open this fall. “It’s not rocket science, but there is a significant level of research necessary and there’s a kind of tedium to it,” he says.

Architects also have to interpret what their clients mean when they ask for certain features. In the case of the Gates and Hillman Centers, Scogin says, there was a strong desire for the buildings to embody the culture of the School of Computer Science and the university. “You have to understand how computer scientists interact together, how they think, how they study,” he says, “and then try to weave that understanding into some sort of a physical response.”

That research received a big boost from the preliminary work done by the SCS program committee. Beginning in September 2004, the committee chaired by computer science professor Guy Blelloch surveyed more than 300 faculty and staff members and students to determine not just their space needs but their quality of life concerns. They visited more than 20 other academic buildings and set short- and long-term goals for the new complex. One of their main goals was promoting interaction and collaboration between faculty, students, researchers, visitors and members of the Carnegie Mellon community outside of SCS.

“These people were visionaries in terms of their goals and expectations,” Scogin says. “If these buildings are successful, it’s because of them.”

Ralph Horgan, Carnegie Mellon’s associate vice provost for campus design and facility development, says the Gates and Hillman Centers as designed should answer all of the requests made by the university and the SCS facilities committee. “The degree to which the buildings adhere to the original intent is enormous,” Horgan says. “Even the more subjective things we asked for were met or exceeded.”

Because of Carnegie Mellon’s commitment to green practices in new construction, keeping the buildings environmentally sensitive was an important objective—but keeping them functional and comfortable was equally important. Scogin’s firm designed the buildings to the planning committee’s checklist and then compared the resulting specifications to the Leadership in Energy and Environmental Design (LEED) standards of the U.S. Green Building Council. The design team was aiming for a silver LEED certification, but will be within range of a gold award, Horgan says. Materials and methods were chosen specifically for their low negative impact on the environment, like the gray water system that will collect rain and melted snow and use it to flush toilets.

The construction process has employed environmentally friendly methods, too. All construction waste is being sorted and separated, and between 50 and 75 percent has been recycled, not sent to landfills. Even the rock excavated for the foundations was reused—some 40 tons of stone were taken down to Panther Hollow, crushed into gravel, and then brought back up for use as construction material. “It was a wonderful way to

→ Creating the Gates and Hillman Centers

preserve what was there,” says director of construction Joe Greenaway.

All of the technologies being deployed in the Gates and Hillman Centers are efficient but proven, Horgan says. “We were not interested in seeking ‘bleeding-edge’ technology,” he says. “We want very comfortable spaces that will work for a long time. We’re trying to balance effectiveness with efficiency.” Many of the energy-saving technologies deployed in the Gates and Hillman Centers are passive (no moving parts) to cut down on mechanical and electronic failures.

Others features of the building may seem to look to the past for inspiration. Until recently, most “modern” office buildings were climate-controlled and hermetically sealed. Unfortunately, that left some occupants too hot, others too cold and everyone breathing stale air. In the Gates and Hillman Centers, all 310 offices are getting a window that opens to admit fresh air. Ventilation systems will draw additional fresh air from outdoors and exhaust the stale air; rotary heat exchangers called enthalpy wheels will use the exhausted air to help raise or lower the tempera-

ture of the incoming air as appropriate without requiring additional energy. And people using classrooms, offices and labs will be able to control the volume and temperature of the air being delivered to every space. (Motion detectors will close air ducts when rooms are unoccupied.)

The desire to make every office “a room with a view” dictated the unusual shape of the buildings, with floor plans and exterior walls that vary from story to story. The resulting design is unconventional, Scogin says, but it fits into the landscape despite the steeply sloping terrain while allowing the complex to have multiple entrances facing Forbes Avenue, the Cut and Wean Hall. Besides, a straightforward “brick box” with 310 offices would have been foreboding—up to 14 stories tall, he says.

In many ways, the Gates and Hillman Centers are coping with the problems—uneven terrain, tight spaces, multiple groups of users—that Henry Hornbostel faced when he designed the original Carnegie Tech campus in the 1900s. Scogin, who has admired Hornbostel’s work both at Carnegie Mellon and at Atlanta’s Emory University, says

his buildings may seem traditional but are far from classical architecture. At Carnegie Tech, Hornbostel disguised the smokestack of what’s now called Hamerschlag Hall with a temple-like rotunda and put sloping corridors inside Baker and Doherty halls to conform to Pittsburgh’s steep slopes. Those ideas were as unusual in their day, Scogin says, as the exterior of the Gates and Hillman Centers may seem to some eyes today.

By putting the power plant at one end of campus and the art school at the other end, Hornbostel also forced engineers and artists, scientists and craftsmen, to share ideas and interact with each other every day. With its bridges to the Purnell Center for the Arts, its proximity to the Heinz College, and its creation of a new east-west corridor on the Carnegie Mellon campus, Scogin hopes the Gates and Hillman Centers will encourage the same kinds of interdisciplinary interactions for years to come. “That kind of dialogue between these seemingly disparate conditions has been what Carnegie Mellon has been about throughout its entire history,” he says.

— Jason Togyer (HS’96)

Giving Back

Scott Dietzen (S’84, CS’88, ’92) immortalizes coworkers with classroom in Gates Center



When asked about some of the successful projects he’s worked on, Scott Dietzen (S’84, CS’88, ’92) borrows an old joke from his days at Transarc Corp. and claims that each team he worked with “inadvertently

lowered its average IQ by letting me join up.”

Scott is being modest. In a productive career, he’s been principal technologist for Transarc, a Pittsburgh-based software company that’s now part of IBM; chief technology officer at BEA Systems, overseeing the WebLogic product family; and president and CTO of Zimbra Inc., developers of a suite of collaborative software products, until that company was acquired by Yahoo! in 2007. Since October 2007, he’s been senior vice president of global communications products for Yahoo!

Scott credits much of his success to the gifted people with whom he collaborated at Transarc, WebLogic, Zimbra, Yahoo! and Carnegie Mellon.

Now he’s giving back by donating a classroom in the new Gates Center for Computer Science, part of the new Gates and Hillman buildings slated to open this fall.

His gift—to be known as the Zimbra Classroom—was made in honor of the team he worked with at the company. Inside that classroom, student innovators will work together to push the envelope of computing—and maybe crack a smile over the fact that a small Microsoft competitor managed to carve out a space inside the Gates Center.

Scott’s gift is an investment in the things that make Carnegie Mellon and the School of Computer Science such innovative environments: our tradition of collaboration, our inter-

disciplinary approach to solving problems, our creativity and our dedication to new ways of thinking. Thanks to the generosity of donors like Scott Dietzen, the next generation of computer scientists will thrive at the university.

Carnegie Mellon recently launched the public phase of its \$1 billion development campaign, called “Inspire Innovation.” As of January we had already raised \$565 million, including approximately \$40 million for the Gates and Hillman buildings.

If you have questions about how you can help the School of Computer Science, whether it’s through scholarships and fellowships, faculty support or gifts toward the Gates and Hillman buildings, please contact me at mdorgan@cmu.edu or 412-268-8576. You can also learn more about the campaign or make your gift online today at www.cmu.edu/campaign.

—Mark Dorgan is principal giving advisor and development liaison for the School of Computer Science.

Apples and Oranges

Two research groups take robotics down on the farm



Ben Grocholsky, project scientist at the Robotics Institute, prepares for another day of research “down on the farm”—in this case, Soergel’s Orchards north of Pittsburgh.

SANJIV SINGH PHOTO

You don’t expect to find Carnegie Mellon computer scientists working in apple orchards and orange groves. But maybe you should, says Sanjiv Singh, a research professor in the Robotics Institute. “I am very interested in automation in agriculture,” he says. “I have gravitated to farm applications because I can see a benefit for society.”

Last November, the U.S. Department of Agriculture awarded \$10 million to two research groups at Carnegie Mellon to develop systems to monitor crop health, spray water and pesticides, and perform certain automated tasks. One team headed by Singh (CS’90, ’95) received \$6 million to develop equipment for apple orchards, while another directed by Tony Stentz (S’84, CS’90) and Herman Herman (CS’93, ’96) of the Robotics Institute’s National Robotics Engineering Center received \$4 million to create technology for use in orange groves.

The award is part of a USDA grant program called the Specialty Crop Research Initiative, which encourages researchers to design systems to improve the growing of plants that are valuable but difficult to cultivate, including fruits, vegetables and nuts. According to the USDA, research into technology used in orchards and groves

hasn’t kept pace with advances in the cultivation of bulkier cash crops such as wheat or cotton.

Some people might wonder why researchers from Carnegie Mellon—an urban university not typically associated with agriculture—are working on farming technology. In fact, the Field Robotics Center has already helped design robots to harvest so-called broad-land crops like alfalfa, and researchers have worked on other projects with agricultural scientists from Penn State, Purdue, Oregon State and Washington State.

“We’re not plant scientists—that’s not what we do,” Singh says. “But we do have people here who are interested in making an impact in agriculture. Robotics is an interdisciplinary science and agriculture is an area of application.”

Many people assume that farm robots would be useful picking apples and oranges from the boughs of trees. Singh says that humans are better at picking produce. Instead, robots can help with things like orchard management. Growers, for instance, don’t always have accurate ways to predict how many apples or oranges they’ll have until the growing season ends, but autonomous farm vehicles could drive through orchards and groves and count the ripening fruit.

Last fall, Singh and his colleagues rode an all-terrain vehicle packed with electronic gear around the grounds of Soergel’s Orchards in Franklin Park, Pa., north of Pittsburgh. As the ATV chugged through rows of apple trees, researchers collected a range of data about the terrain and land features. The data will help them develop algorithms to control autonomous guidance and obstacle avoidance systems for use in orchards.

Singh and his team hope to create sensors and GPS-guided navigation systems that enable robots to independently monitor orchards like Soergel’s. Some of their sensors might be able to detect chlorophyll levels—a sign of plant health—and could alert growers to potential problems such as diseases or bug infestations.

“Big producers care a lot about the crop they’ll get,” Singh says. “The amount of fruit varies from season to season because of weather, insects and

“Big producers care a lot about the crop they’ll get. The amount of fruit varies from season to season because of weather, insects and temperature fluctuations. Producers want to know how much to expect and typically they do it by looking at the trees before harvest. This could help them see crop yield ahead of time.”

Sanjiv Singh (CS’90, ’95)

temperature fluctuations. Producers want to know how much to expect and typically they do it by looking at the trees before harvest. This could help them see crop yield ahead of time.”

Such robots could also cut grass, trim trees or spray for pests as they roam. Singh doesn’t know whether the robots his team develops will do several things at a time or specialize in certain tasks. But he’s happy to show another way that Carnegie Mellon roboticists are developing applications for use outside of heavy industry and defense. “I think this is a fantastic way of contributing to society,” he says.

—Meghan Holohan is a Pittsburgh-based freelance writer who writes frequently for MentalFloss.com and whose work has appeared in magazines such as Geek Monthly.



DAVID BACHMAN PHOTO

Celebrating the creation of the McWilliams Center for Cosmology are Carnegie Mellon President Jared Cohon and wife Maureen Cohon; Bruce (S'78, '81) and Astrid McWilliams; and Barbara Gilman and her husband, Mellon College of Science Dean and Buhl Professor of Theoretical Physics Fred Gilman.

The Stars Align

Center unites cosmologists, computer scientists

Carnegie Mellon Provost Mark Kamlet likes to say that he would have more respect for physicists if they could find out where they misplaced 95 percent of the universe. Behind his joke is a serious question—maybe the most fundamental question facing science today. The vast majority of the universe is made up of dark matter and dark energy, about which very little is known. Future discoveries about the nature and origin of the universe hinge on learning more about these mysterious substances.

The new Astrid and Bruce McWilliams Center for Cosmology at Carnegie Mellon will be probing the mysteries of the universe while tackling another weighty problem in computer science—namely, how to process all the information produced by scientific research. Take, for example, the Large Synoptic Survey Telescope being constructed on a mountain in northern Chile. Carnegie Mellon is one of 18 American universities collaborating on the LSST, which will soon spend 10 years taking high-resolution images of the entire visible night sky using the largest digital camera ever created. The LSST is eventually expected to take more than 200,000 images every year, generating 100 terabytes of data every week.

“In almost all of the physical sciences, we’re seeing gigantic amounts of data being gener-

ated,” says Peter Lee, head of the Department of Computer Science. “There’s just no way humans can sift through it all.”

Carnegie Mellon alumnus and trustee Bruce McWilliams (S’78, ’81) recognized that cosmology was complemented by computer science in many ways. For instance, the models that SCS researchers have created to simulate air currents and water flow are similar to the fluid models that cosmologists use to simulate the formation of galaxies and black holes, says Adrien Treuille, assistant professor in the computer graphics group at the Robotics Institute. “If interactions generally unfold in predictable ways, we can use far fewer variables to represent the system and have correspondingly faster simulations,” he says. And cosmologists’ need to process large digital images and create complex graphics to depict physical phenomena requires advanced methods of computer modeling, data mining, storage and retrieval.

The founder and CEO of Tessera Technologies Inc., a San Jose, Calif., based company that develops miniaturized integrated circuits and optical sensors, McWilliams decided that Carnegie Mellon’s heritage of interdisciplinary research made it an ideal place to work on cosmic-scale simulations. Last April, he and his wife created the McWilliams Center as a collaborative effort between the Mellon College of Science and SCS. Lee says McWilliams has been engaged first-hand in creation of the center “at every step along the way.”

Cosmic-scale simulations require an enormous amount of processing power. A new high-performance computer cluster donated by the Betty and Gordon Moore Foundation (and dubbed “the Moore Machine”) will be shared by the Computer Science Department and the McWilliams Center and will be devoted solely to research on large-scale simulations. Additional computational oomph is coming from the Pittsburgh Supercomputing Center, home of a cluster of super-fast computers nicknamed “Ferrari.”

Lee sees a role for computing in cosmology beyond data crunching. A growing field called “eScience” aims to use machine learning to automatically discover scientific facts. As computers process the massive amounts of data coming from telescopes, for example, they might be looking for galaxies of a certain shape. But they’ll be seeing a lot of other things, too—including potentially valuable new discoveries.

“Can we make computers at this scale intelligent enough to know when there’s something special?” Lee asks. “Can we automate discoveries?”

Isaac Asimov said the most exciting phrase in science wasn’t “eureka!” but “that’s funny.” If computers can be taught how to spot “funny” anomalies in data, Lee envisions a day when they’re making discoveries not only about galaxies and dark matter, but also spotting things like asteroids heading toward Earth.

Because the McWilliams Center combines both pure scientific research with the possible discovery of “practical things” like asteroids on collision courses, Lee says it’s well positioned to attract funding from programs such as the LSST and the National Science Foundation’s Cyber-Enabled Discovery and Innovation program. At any rate, the pace of discovery won’t be slowing down any time soon.

—Karen Hoffmann (S’04) is a Pittsburgh-based freelance writer and photographer who covers science, health and the environment.

Steady and Sure

Scarab rover prototype passes early trials in Hawaii

The creature crawls on the edge of a dormant volcano called Mauna Kea in Hawaii. As it rolls along, its chassis adjusts to the terrain, moving up and down to help the tires grip the black sandy surface. This beast resembles a bumper car on dirt-bike wheels, but it's no toy. Known as Scarab—because its hull is shaped like a beetle's body—this autonomous robot is a prototype for a lunar rover.

Onboard lasers scan the landscape and plot a map that looks like the 3D topography of a video game. The map will help Scarab find the best route down the steep hillside.

From a nearby base camp, associate research professor David Wettergreen (S'87, E'89, CS'95) and his team watch Scarab in action. He wants Scarab to descend the hill using a drainage ditch—the easiest way down the slope. But Wettergreen has to be patient as he observes the rover. It takes eight hours for Scarab to cover one kilometer.

It's critical that Scarab makes the right choice. NASA wants a lunar rover that can explore craters at the poles of the moon without human direction. Wettergreen and his team are studying how Scarab will climb slopes and drive through loose volcanic soil. If Scarab doesn't find the easiest path, they'll have to rework the robot and its navigation software.

Scarab slowly turns and heads down the drainage ditch. Relief washes over Wettergreen and his colleagues. The rover made the right choice. As the robot performs test after test, Scarab capably avoids obstacles such as rocks. Instead of going around all of them, Scarab sometimes lifts up its body so the rocks slide underneath.

To support human exploration and missions to other planets, NASA intends to create a base on the moon. Before that can happen, the space agency needs vehicles that can explore parts of the moon that have never been seen by humans. Scientists think craters at the poles of the moon have frozen water in them and NASA wants a rover that can drill and sample the dirt. That's why NASA scientists approached Wettergreen and Red Whittaker, Fredkin Professor of Robotics and director of the Field Robotics Center, who each have extensive experience creating rover prototypes.

Scarab is able to crawl in and out of craters with sides as steep as 29 degrees. It can crawl over a surface until it arrives at a drill site. It then lowers its belly onto the soil to stabilize the rover as it drills into the surface and extracts a core sample of the material. On board Scarab is a drill developed by the Canadian Space Agency and an automated chemistry lab designed and built by NASA. After the drill extracts “regolith”—the loose combination of dirt that covers solid rock on the moon—it puts a sample in a crusher, which pulverizes it to a fine powder. Then it places the powder in an oven that heats it to 900-degrees Celsius, releasing all of the volatile materials as gases and creating a sintered ash. From a chemical analysis of these gases, Scarab's internal lab can determine how much of each

In October and November, Wettergreen and his team put Scarab through a battery of experiments on Mauna Kea, testing the robot's mobility and a new set of lunar wheels made by Michelin for exclusive use on the moon. When exposed to extremely cold temperatures—like those on Earth's moon—rubber tires will become brittle and crack. Scarab's special wheels were made of fiberglass and were designed to withstand low temperatures without compromising mobility.

“It was the first time the wheels were tested outside of the lab and over long distances,” Wettergreen says. However, as Scarab rolled over the ground again and again, the tread wore down and lost traction. To counteract this, Wettergreen and his team crafted “grousers”—new tread-like ridges of metal—to replace the worn-down tread on the



David Wettergreen, associate research professor at the Robotics Institute, demonstrates Scarab's ability to lift itself over obstacles. The rover is testing concepts that could be deployed in a future lunar vehicle.

chemical compound is in the ground. If there were water frozen below the surface of the moon, the rover's computer would be able to tell.

During testing in Hawaii, Scarab drilled five holes, each time taking anywhere from one to three hours. The processing of the rock took another eight to 20 hours.

Scarab's body was built by Fringe, one of the Carnegie Mellon student organizations that enter the annual Sweepstakes (or “buggy”) races during spring carnival. It is made of carbon fiber and shaped so that it won't snag obstacles and can support the weight of the robot when it lowers to the ground.

wheels. These helped the wheels grip the soil, but they required more power to drive the rover. Without the grousers, the researchers discovered the lunar tires had 20 percent less traction but also used 50 percent less energy than rubber tires. Using less power is key, because it allows more of Scarab's energy to be devoted to other tasks.

“The object was to demonstrate the complete process of lunar prospecting,” Wettergreen says. “We're developing models of vehicle performance that are calibrated well enough to predict how (such a rover) would work on the moon.”

—Meghan Holohan

Saying 'Hala' to Hala

Qatar welcomes world's third roboceptionist

When people enter the new home of Carnegie Mellon Qatar in Doha's Education City, they're often overwhelmed by the structure's beauty. And coming this year, they'll also be greeted with "hello" or "marhaba" from a robot named Hala.

"Hala," whose name means "welcome" in all Arabic dialects, is the new "roboceptionist" in Education City. A "sister" robot to "Tank," who debuted at Newell-Simon Hall on the Pittsburgh campus in September 2005, Hala is part of a new project sponsored by the Qatar National Research Fund and Qatar's Undergraduate Research Experience Program to investigate human-robot interaction in a mixed cultural setting.

She's able to answer questions in English and Arabic about both Education City and Carnegie Mellon's home campus in Pittsburgh. Brett Browning, a senior systems scientist with the Robotics Institute in Qatar, says the goal was to devise robots that can interact with visitors in culturally appropriate ways. Using cameras and pattern recognition software, Hala can detect whether an approaching guest is male or female. She's also able to tell if someone is wearing traditional Gulf dress or Western clothing. That helps to determine whether she initiates a conversation in English or Arabic. Because gestures and facial expressions carry different meanings to Arabic and English speakers, Hala is also tuned to react in culturally sensitive ways, Browning says.

Hala is being developed by an intercampus, multinational team. In Pittsburgh, the team includes Reid Simmons—research professor of robotics and developer of Carnegie Mellon's first roboceptionist in 2004—and robotics doctoral student Maxim Makatchev. The Qatar team includes associate teaching professor Majd Sakr; Browning; and research staff Ameer Abdulsalaam, Imran Fanaswala and Wael Al Ghazzawi. Undergraduate students Hatem Alismail (CS'09), Keghani Kouzoujian (CS'09) and Wadha Al-Adgham (CS'09) also contributed.

Alismail and Kouzoujian went to Pittsburgh in the summer of 2008 to work alongside the creators of Tank. "I wanted to learn more about robots and get more involved in research," Kouzoujian says. "The roboceptionist allows us to put things into action in a way that will have a lasting impact on campus. And it's pretty cool, too."

Hala has a pan-Arabic personality—her gestures reflect a variety of middle Eastern influences instead of one specific nation or culture—and speaks without a local dialect. The team made her female since most students at Education City are women. Although Hala isn't mobile, Browning envisions the day when she might have wheels that allow her to move freely around the building. "Not only could she tell people where they needed to go, she could actually show them," he says.

Browning also hopes the Hala project and the monetary support from the Qatar Foundation will entice other students to join the culture of research being cultivated in Qatar. "This is just the start of introducing robots to the Gulf," he says. "Qatar is a society that is transforming itself in an unprecedented way. It's a great place for robots and artificial intelligence to have a huge impact."

—Andrea L. Zrimsek is a writer at Carnegie Mellon University in Qatar.



ANDREA L. ZRIMSEK PHOTO

Hatem Alismail (CS'09) performs some preliminary work on Hala, the new roboceptionist at Carnegie Mellon Qatar.

Bugfix

"The Lasting Legacy" (The Link, Fall 2008) misspelled the name of Jacobo Carrasquel, associate teaching professor in the Computer Science Department.

The Link corrects all known errors of fact. Please send corrections to TheLink@cs.cmu.edu or write to The Link Magazine, Office of the Dean, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213.



JOHN BARNA PHOTO

Bruce Nelson Professor
of Computer Science

Manuel Blum

Manuel Blum came to Carnegie Mellon as a visiting professor in 1999 and has been the Bruce Nelson Professor of Computer Science since 2001. The 1995 winner of the A.M. Turing Award, he is a member of the National Academy of Sciences and the National Academy of Engineering and a fellow of the American Academy of Arts and Sciences. Husband of Lenore Blum, Distinguished Career Professor of Computer Science, they are together the parents of a third member of the SCS faculty, Computer Science Professor Avrim Blum. Manuel Blum recently spoke to Joanna Steward, founding managing editor of *The Link*, about problem-solving strategies and why Paul Erdős was correct when he said that mathematicians are machines that turn coffee into theorems.

Solving Hard Problems

I'd like to understand how to create high-level plans or strategies for solving hard mathematics and computer science problems. But what does it mean to devise a strategy? How do humans create strategies? How can we get machines to create them?

When you have a really hard problem and you've studied the literature, thought about the problem and looked in many different directions for a

solution, and none of the ideas work, where can you find new ideas? One way to create a strategy where none exists is to take the problem that you want to solve and modify it. Simplify it or change it in some way so that it becomes easier to solve. Solve the simpler problem and then use that solution as a kind of template—a strategy—for solving the original problem.

For example, in Avrim's algorithm design class (15-451), which I greatly enjoy teaching, the TAs help me to modify each week's homework problems. That way the students can't just look up answers. It's important to keep the guts of the problems, and also to create problems that are new. We talk over the changes—what the TAs decided to modify, whether the changes are instructive, whether the problem has lost or gained something important, and so on. Basically, I get to try out some of my informal ideas about strategies and how to go about changing problems to make them easier or different.

Turning Coffee into Theorems

One of the things we humans learn is that when you're working out a hard problem and you aren't getting anywhere, it's a good idea to take a break—go out and get a cup of coffee. When you come back, hopefully refreshed, you can often come up with some idea you didn't have before.

But how do you turn advice like that into advice for a computer? You can't tell a computer to take a coffee break. So what's going on in our minds when we take that coffee break? Part of it is that when we come back, we are not back at exactly the same point where we were before. We start fresh in a new state. It's the sort of thing that humans do ... and computers can be made to do it, too.

Rip Van Winkle

I remember when I was young thinking, "Wouldn't it be neat if we could just go to sleep for a hundred years and wake up and see what the world was like?" Wouldn't it be neat to see the progress that had been made?

I feel like I'm living that right now—I don't have to go to sleep! The world is changing so quickly. Advances in science—in health, biology, computer science, physics, all of these things—are moving so quickly that we're constantly faced with a new world.

In the 19th century, we got used to the fact that machines could do hard physical labor better than humans. In the 20th century, computers became better than humans at memorization and information retrieval. Now we're in the 21st century. I don't know exactly how it's going to work out, but I personally believe that computers are going to be able to do any intellectual activity that humans can do. It's something I'm looking forward to seeing.



Christina Levkus, math teacher at Steel Valley High School in Munhall, Pa., helps a student work through a geometry lesson on an intelligent tutoring program. The classroom is one of the “LearnLabs” collecting educational data for the Pittsburgh Science of Learning Center.

WADE H. MASSIE PHOTO

Head of the Class

By Jason Togyer (HS'96)

The Pittsburgh Science of Learning Center—a joint effort of Carnegie Mellon and the University of Pittsburgh—is building the world’s largest repository of educational data collected directly from classrooms. Mining that information is providing new insight into the ways that children and adults learn.

It’s a Wednesday morning, and in Christina Levkus’s classroom at Steel Valley High School outside Pittsburgh, a group of 10th graders is learning how to derive the measures of the angles in a triangle using a computerized tutor. Green-colored bars, or “skillo-meters” (think “thermometers for skill”) show how much a student has learned in any given lesson. When an indicator turns gold, the student has mastered that skill.

Suddenly, one boy cries out: “This is crazy! Why can’t I get any gold bars?”

The stricken sophomore glances at the computer next to his. “No fair,” he says to the classmate working on that machine. “How come you’ve got gold bars?”

“Hey!” she says, pointing at the screen. “I can’t get this one stupid bar to move at all.” As they commiserate, Levkus walks over to see why they’re struggling.

It’s a typical scene in any of the 2,600 schools that use Cognitive Tutors, the interactive, intelligent tutoring programs developed at Carnegie Mellon in the 1990s and marketed by Pittsburgh’s Carnegie Learning Inc. The difference between other classrooms and this one is that Levkus’ students are using a course that’s part of LearnLab, a nationwide resource for education research created by the Pittsburgh Science of Learning Center, or PSLC, a joint venture between Carnegie Mellon and the University of Pittsburgh and sponsored by the National Science Foundation.

The course that Levkus’ students are using is capable of measuring many variables. Data are being collected for a study by Vincent Alevan, assistant professor in the Human-Computer Interaction Institute, and post-doctoral fellows Ryan Baker (CS’05) and Ron Salden. As Levkus’ students work through the lessons, their mistakes and right answers are logged on a server. The software is also

counting the minutes it takes for students to work through problem sets and the number of times they ask for online help.

Each point of information collected is stored in the “DataShop,” a massive repository of educational data created and maintained by the PSLC. Every day, DataShop is receiving similar data from schools and universities using six course-length tutoring systems in algebra, geometry, physics, chemistry, Chinese and English; as well as dozens of other learning environments, some created specifically for certain experiments. So far this year, more than 14,000 students have been part of LearnLab experiments, ranging from elementary, middle and high schools in Pennsylvania, New Jersey and Florida to university campuses in Hawaii, British Columbia, Denmark and Germany.

As a result, DataShop is now arguably the world’s largest public repository of empirical educational data ever gathered in the field, with a corpus of approximately 100,000 hours of student instruction that comprises 22 million different transactions between students and tutoring programs.

“We really have established that intelligent tutoring systems can be a great way not just of delivering instruction,” says Ken Koedinger (HS’88, ’90), a professor of human-computer interaction and psychology at Carnegie Mellon, “but that they also provide the technical infrastructure for helping us understand human learning.”

Koedinger is codirector of PSLC with Charles Perfetti, research director and senior scientist at the University of Pittsburgh’s Learning Research and Development Center, or LRDC. Members of the executive committee include Alevan; Maxine Eskenazi (HS’73), associate teaching professor in the Language Technologies Institute (see sidebar); David Klahr (TPR’65, ’68) of Carnegie Mellon’s psychology department; Marsha Lovett (HS’91, ’94) of Carnegie Mellon’s Eberly Center for Teaching Excellence; and Julie Fiez, Tim Nokes and Lauren Resnick of the Pitt psychology department. Michael Bett (S’86, TPR’00) is managing director.

Created five years ago with \$25 million from the National Science Foundation, PSLC learned in February that its work would be funded for another five years with a similar grant.

Perfetti, University Professor of Psychology at Pitt, serves as chief scientist of PSLC. “It’s really hard to overemphasize the value” of the LearnLab infrastructure, he says. Classroom research before the creation of LearnLab required high levels of effort and repeated trips into the field, Perfetti says, likening them to visits to the dentist. “Every time you did one, you wondered if it was the last study,” he says, adding that the more data the researcher collected, the harder they were to analyze.

But converting a classroom that’s already using Cognitive Tutors into a LearnLab is a relatively simple process. And because so many schools are already using Carnegie Learning products, >>>

“We really have established that intelligent tutoring systems can be a great way not just of delivering instruction, but that they also provide the technical infrastructure for helping us understand human learning.”

Ken Koedinger (HS’88, ’90)



Ken Koedinger

they're familiar with Cognitive Tutors and comfortable with Carnegie Mellon staff. That's a built-in network PSLC research manager Gail Kusbit uses whenever a researcher needs to conduct an experiment. "We're very fortunate in that so many schools say yes," she says.

As a result, LearnLab and DataShop are the foundation of a network for continuous data collection that Perfetti envisions becoming an educational research resource "for the world." With LearnLab, he says, education and cognitive psychology researchers "never have to think twice" before committing to a classroom study. "You just do it."

LearnLab is only one important initiative of PSLC's researchers. Others include the Cognitive Tutor Authoring Tools, or CTAT, a software suite that allows non-programmers to design their own intelligent tutors; TuTalk, which enables creators of intelligent tutors to create less obtrusive, more intuitive online help programs; and TagHelper, a suite of programs that partially automates the process of annotating documents in Chinese, English, German and Spanish, making it easier to classify texts gathered in protocol analysis ("think-aloud" or dialogue data).

PSLC's express purpose is to encourage experimental research into "robust learning"—giving students a deeper appreciation of the material and the ability to transfer their knowledge into new situations. Studies indicate that students who learn "robustly" in a subject area find it easier to master new skills in the future.

Identifying "robust" learning strategies after a class ends is difficult. Assessment tests provide a crude measurement of how much a student learned, but

don't offer any insight into why a particular student did (or didn't) retain knowledge, and don't document struggles and milestones in the learning process. Conventional wisdom, for instance, holds that "practice makes perfect," and that students learn more if they have lots of homework. Research through LearnLab indicates that isn't necessarily so.

"Problem sets are good, and you need practice, but especially for students just beginning a topic, it's not enough," Koedinger says. Instead of drilling students on homework problems, researchers have had them use intelligent tutors that alternated solved examples with new problems. As a result, the students retained more information and did better on tests. "They spent more of their brain power on understanding rather than just getting the problems done," Koedinger says.

PSLC combines Carnegie Mellon's long heritage of research into cognitive psychology and computer science—a tradition reaching back to Herbert Simon and Allen Newell—with the University of Pittsburgh's School of Education and Learning Research Development Center. Founded in 1963, the LRDC includes research in cognitive science, developmental and social psychology, organizational behavior and education policy to understand the ways children and adults learn.

"PSLC is the kind of thing that shows off these intellectual resources that we have in Pittsburgh that make it such an exciting place," says Carolyn Penstein Rosé (HS'94, CS'97), assistant professor in Carnegie Mellon's Language Technologies Institute and Human-Computer Interaction Institute. Rosé is collaborating with Resnick, former director of the LRDC and Distinguished University Professor of Psychology and Cognitive Science at Pitt, on research into social and communicative factors in learning. They're examining how classroom conversation and interaction—including online chats and collaboration over the Internet—enhance the ways students master new material. Their goal is to teach students how to articulate their reasoning, and recognize when other students are also reasoning through a problem.

Although she's long admired Resnick's research, Rosé hadn't found a way to work with her until PSLC provided a network for collaboration. "PSLC encourages an active exchange between different research communities that might not have gotten involved together otherwise," Rosé says. "It's really evolved organically."

Through face-to-face meetings and a lively wiki where users propose, document and debate theories, PSLC unites researchers in education, psychology, machine learning, human computer interaction, language technologies and other fields. Many PSLC activities center on experi-

ments conducted via LearnLab courses, or by mining data from the DataShop.

Probing data logged by geometry Cognitive Tutors, for instance, Alevin and other PSLC investigators noticed that a higher than expected percentage of students preferred to make repeated errors rather than ask for online help. Some students were worried (incorrectly) that their "skillometers" would go down. Others weren't sure when to seek assistance. Alevin and his colleagues built a tutor that worked alongside the geometry program to teach students when to ask for help. The results were mixed, he says—students got better at spotting hypothetical situations when they should look for assistance but weren't putting that information into practice.

Still, the experience suggested important new areas for further research into motivational techniques and the role of positive and negative reinforcement, Alevin says. And a recent analysis suggests that students exposed to the "help tutor" used smarter strategies on subsequent exercises after that agent was removed, he says. "We're very excited about this result," Alevin says. "To the best of our knowledge, it's the first time that an intelligent software tutor has been shown to have a lasting effect on student learning at the meta-cognitive level."

After all, the ultimate goal of the PSLC's research isn't just to improve student skills in math or science, Alevin says: "We're trying to help students become better learners."

Humans have been learning new skills since before the dawn of recorded history. But convincing educators to examine the ways students learn—and the ways teachers should teach—is surprisingly difficult, says Klahr, Walter van Dyke Bingham Professor of Cognitive Development and Education Sciences at Carnegie Mellon, who has been studying elementary and middle school science education for more than a decade.

Many science teachers abandon the scientific method when they're evaluating the success of their own lessons, Klahr says. "They love their subject matter and believe they can teach it, but they fail to apply the same rigorous procedures to evaluating the impact of their teaching methods as they would apply to their own scientific domains."

The problem is that educators spend years in the classroom as students, Klahr says, and come out with strong beliefs about what works and what doesn't. Changing their minds requires rigorous, verifiable data obtained through controlled experiments, and gathering that data isn't easy. A researcher can observe a classroom of 30 students at a time, but can't easily study entire school districts. And a researcher designing an experimental lesson can't ensure that every teacher will deliver that lesson in exactly the same way. >>>

Unraveling Language

Maxine Eskenazi knows how much can depend on the meaning of a single word.

While she was living in France, her French mother-in-law sent one of her American aunts a gift of delicate chiffon fabric.

But the French have a different name for the sheer material—mousseline—while “chiffon” means “rag.” Eskenazi’s mother-in-law received, therefore, a well-intentioned note thanking her for the “nice rags.”

Such incidents, along with Eskenazi’s experiences teaching English in France, instilled in her a lifelong appreciation for the intricacies of language learning.

Learning a new language is hard work. Too many unknown words or a dull subject can make the effort seem unrewarding. Aware of these issues, Eskenazi, an associate teaching professor in the Language Technologies Institute and a member of the PSLC executive committee, and LTI professor Jamie Callan developed software to customize students’ reading materials, helping them engage and learn more rapidly.

The software they created—Reader-Specific Lexical Practice for Improved Reading Comprehension, or REAP—combs the Internet for documents that fit the reading level and interests of the individual student. Finding real documents from the Web was crucial. “I wanted authentic texts for students, not things teachers make up that wouldn’t reflect how language was really used,” Eskenazi says.

REAP first gives a vocabulary test to find words that a student doesn’t know. Because earlier studies found that students were more motivated to learn a language when they were reading about topics that interested them, students are also quizzed about their interests. The software then seeks out stories and articles with words the

student needs to learn in the subject areas they enjoy—and with enough novelty to keep the budding reader challenged but not frustrated. REAP helps to teach vocabulary using context, rather than just definitions.

Teachers also were asked for their input on the technology. “It’s extremely important when you’re creating software that you have your ears wide open to what teachers need,” Eskenazi says. During initial studies, instructors indicated they wanted to be more involved in the learning process, so researchers created an interface teachers can use to find documents for class discussion using specific topics and vocabulary.

REAP is presently in use at Pitt’s English Language Institute, where teachers have noticed students—far from being bored—are copying down URLs so they can go back and peruse the readings at their leisure.

“If we can motivate students to read more, that’s a very good thing,” Eskenazi says. REAP is also available in French, and the team recently received a grant from the Carnegie Mellon-Portugal Initiative to develop a Portuguese version.

Once a student has mastered written vocabulary and grammar, there’s still the daunting prospect of pronunciation, where mistakes are inevitable and can be embarrassing, especially early in the language-learning process. To help with this problem, Eskenazi has created software called NativeAccent. A product of Carnegie Mellon spin-off Carnegie Speech, of which Eskenazi is chief technology officer, the interactive tool allows a student to practice speaking in front of a computer.

Pronunciation errors are corrected using speech-recognition technology. Through a cycle of speaking and receiving corrective feedback, the student gradually builds enough confidence to converse with other people. NativeAccent and other Carnegie Speech products have been used by speakers of more than 40



languages. In January, Carnegie Speech launched a new product specifically for international pilots and air traffic controllers, who must meet new standards for spoken English by 2011.

Eskenazi’s future research will address the fact that students coming from different cultural backgrounds may present different learning challenges. For example, some Arabic-speaking students have an easier time learning to speak than read because they don’t have a strong early exposure to written language. Students from East Asia can have the opposite tendency.

Eskenazi envisions an enhanced version of the REAP software that could gather video instead of text and use speech recognition to teach spoken language. She’s also working on a tool to teach meta-linguistic cultural concepts.

Such technologies will help ensure that—when learning a new language—future students steer clear of too many avoidable faux pas.

—Karen Hoffmann (S’04)

Cognitive Tutors don’t have that problem. “They’re infinitely patient, and you can test things out for as long as you want,” says William Cohen, associate research professor in the Machine Learning Department, who worked on development of SimStudent, part of the CTAT suite. Cognitive Tutors provide researchers with a degree of control over an experiment’s parameters that human teachers can’t duplicate. Because they can be deployed at many sites simultaneously, the amount of data collected is orders of magnitude larger than anything individual researchers can gather independently.

“When you have a weak signal and a lot of noise, you need a lot of data,” Klahr says. “Now you can do a very sophisticated analysis because the data sets are so large.”

Even relatively small data sets, however, are providing useful information. Tim Nokes is an assistant professor of psychology at Pitt who studies physics education. Traditional physics courses, he says, rely heavily on showing students examples of different problems, then expecting students to solve new problems by applying the same formulas. Unfortunately, Nokes says, some of the best students finish a course with a good

understanding of formulas and a weak understanding of physical concepts.

“There’s a real disconnect between the description of the concepts and the examples,” he says. “They tend to see a problem as a ‘pulley problem’ or a ‘lever problem,’ and not a ‘Newton’s Second Law of Motion’ problem.”

Through LearnLab, Nokes conducted an experiment using 78 midshipmen taking a physics course at the U.S. Naval Academy in Annapolis, Md. One group received a standard lesson with a completed example >>>

Charles Perfetti



KEN ANDREYO PHOTO

problem and was asked to solve a similar problem. A second group was asked to examine and explain each step in the sample problem. The third group was given two examples and asked to compare and contrast them. All of the students were then tested on an intelligent tutor equipped as a LearnLab. The students in the second and third groups had better scores, solved the problems more quickly and required less online help, Nokes says, demonstrating they had better mastered the concepts.

Without PSLC and LearnLab, Nokes could have conducted a similar study in a controlled setting like a laboratory, but such experiments lack verisimilitude. They don't take into account variables like a student's motivation to do well in a certain course. "It's not always clear to me as a researcher that something that works in a lab will work in a classroom," he says.

Experiences like that prove the value of the LearnLab concept, says Marsha Lovett, associate director of faculty development in Carnegie Mellon's Eberly Center for Teaching Excellence and associate research professor of psychology. "This is intensive, moment-to-moment,

"This is intensive, moment-to-moment, click-stream data, not just 10 minutes in a lab."

Marsha Lovett (HS'91, '94)

click-stream data, not just 10 minutes in a lab," says Lovett, coprincipal investigator on a project to create an intelligent statistics tutor. "And you can collect from thousands of students a semester's worth of data each."

Not all of the studies being run through the PSLC require collection of new data, says Ryan Baker, a post-doctoral fellow in the HCII. Many experiments can be performed using DataShop's existing corpus—some of which was gathered back in 1996, before the creation of PSLC, but which has been transferred into the system.

"If you're doing educational data mining, (the PSLC) is the place to be," says Baker, technical director of DataShop and associate editor of the Journal of Educational Data Mining. "Educational data mining is a great new frontier," he says, because it's forcing computer scientists to develop entirely new strategies and algorithms. Many existing data-mining tools look for patterns, word associations or clusters but don't account for hierarchy and variables that depend on each other, Baker says. For instance, he says, if a particular student asks for help, researchers need to be able to interpret that request in light of the same student's previous requests for help.

PSLC has developed applications to plot learning curves, error rates and other measurements for DataShop users, and raw data can also be exported in tab-delimited format. But the amount of data has doubled in the past year and will probably increase a hundredfold in the next two years, Baker says. As a result, scalability is becoming a concern; one fast-moving research area in the PSLC is development of more powerful data-mining tools.

If there's a downside to the LearnLab infrastructure and DataShop, it's that not enough researchers outside of Pitt and Carnegie Mellon are using it. Klahr (quoting a favorite Koedinger analogy) says it's a classic "toothbrush problem"—researchers treat other peoples' theories like toothbrushes, and they don't want to use someone else's.

If the same "not-invented-here" bias also applies to educational data, it may finally be weakening, Baker says. DataShop currently has more than 450 users in a dozen countries, and at the first annual International Conference on Educational Data-

Mining, held last June in Montreal, Quebec, nine of 17 papers included material from DataShop. "The educational data-mining community does see us as a resource, and they're getting excited about it," he says.

Baker is developing software that spots when students using a Cognitive Tutor are getting bored and "gaming the system"—solving problems by guessing or abusing online help instead of reasoning through the steps. "I believe that the PSLC, five years from now, will have made enormous progress toward predicting how an individual student will respond to certain methods of instruction," he says.

Lovett looks forward to the day when intelligent tutors are able to provide smarter feedback to students by analyzing a semester's worth of data. A future intelligent tutor might spot students who cram the night before tests and do poorly, she says. It could then suggest smarter strategies for learning. "If you have all of this learning data coming in moment-by-moment, then why not feed it back to the instructors and the students themselves in a way that's meaningful?" Lovett says.

There are many innovations to come, Koedinger says, and much to be discovered about the way human beings learn. So little experimental data has been gathered from classrooms that it's a vast territory waiting for exploration, he says.

"It's also a particularly challenging one, because there are so many political, ethical and social issues that float around," Koedinger says, "but in the science of learning, there are really still surprises to be had."

—Jason Togyer (HS'96) is managing editor of this magazine. Email him at jt3y@cs.cmu.edu.



Tuomas Sandholm

GLENN BROOKES PHOTO

Winning Hand



“Pokerbots” like the ones Tuomas Sandholm is developing are a likely bet to revolutionize artificial intelligence, just as chess-playing computers did two generations ago.

By Byron Spice There’s high drama at the 2008 World Series of Poker in Las Vegas: Antic poker pro and 11-time champ Phil Hellmuth is all-in, betting his entire stake on this hand. With one card yet to be dealt, he holds the ace of hearts and queen of diamonds in his hand; the king of diamonds and the three, four and 10 of hearts are face up in front of him.

Any number of cards could win the pot for Hellmuth and keep him alive in the competition—any ace, any queen,

either of two jacks, or any heart at all, for heaven’s sake. But the dealer turns over the two of spades. Hellmuth is done, his dream of a 12th championship on hold for another year.

Halfway across the country, at the Association for the Advancement of Artificial Intelligence conference in Chicago, the tension at the third annual Computer Poker Competition is . . . nonexistent. Play ended before the AAAI’s July conference even began and it only



GLENN BROOKES PHOTO

Sam Ganzfried

remains for the competition chair, a computer science Ph.D. student from Carnegie Mellon named Andrew Gilpin, to announce the winners.

“Computer poker is not exactly a spectator sport,” Gilpin says later. That’s an understatement. Working by himself, Gilpin spent four solid weeks running the competition on Carnegie Mellon computers, pitting 15 computer poker programs, or “pokerbots,” from seven countries against one other in more than three million hands of Texas Hold ’Em. As the computers cranked away, he resisted the urge to peek at the results. “I just made sure the programs didn’t crash or didn’t do something malicious and that the results were being generated,” Gilpin says.

What computer poker lacks in drama, however, it more than makes up in intellectual challenge. Poker has taken the place of chess as the game that’s driving artificial intelligence research, giving rise to these annual competitions at the AAAI and, in 2009, at the International Joint Conference on Artificial Intelligence, or IJCAI.

“The 2008 competition was great for us,” says Tuomas Sandholm, professor of computer science

and leader of Carnegie Mellon’s computer poker initiative. For the first time, a Carnegie Mellon pokerbot developed by Sandholm, Gilpin and a collaborator at the University of Munich, Troels Bjerre Sørensen, collected the most chips in Heads-Up Limit Texas Hold ’Em, winning one of the contest’s four categories. Carnegie Mellon also collected the most chips in Heads-Up No Limit Texas Hold ’Em, but that category’s winner was determined by a runoff rather than total chips; the Carnegie Mellon bot ended up in third place.

Winning any contest is nice, but the stakes are particularly high for poker, Sandholm says. The players must make strategic decisions while knowing little about the cards their opponents hold. That contrasts sharply with chess, which offers a dizzying array of possible moves, but also provides complete information about every move that’s made. Adding to the uncertainties of poker is the likelihood that bets placed by opponents can be bluffs.

Making decisions based on incomplete or misleading information is a common dilemma for business and military leaders, so the payoff in developing computer programs that can solve a poker game

is potentially great. “A lot of real-world situations have uncertainty in them, and you have to deal with uncertainty,” says Sandholm, whose research focuses primarily on e-commerce, such as automatically setting rules for auctions or negotiations. Take auctions of radio spectrum by the Federal Communications Commission, for instance. “How do you bid rationally in these auctions?” Sandholm asks. “Nobody knows how. Even if you knew the true value of a piece of the spectrum, bidding at its true value is not an optimal strategy.”

Perhaps a poker-playing algorithm could help. “It may be rational to bid on things you don’t intend to win, just to drive up your competitor’s prices, or to steer them to other parts of the spectrum,” Sandholm says.

Mobile robotics is another realm where uncertainty is a given. On its way to winning the 2007 Defense Advanced Research Projects Agency’s Urban Challenge robot race, Tartan Racing’s robotic SUV, Boss, had to make allowances for vehicles that proceeded out of turn at four-way stops. (See “A November to Remember,” *The Link*, Issue 3.0.) But as autonomous vehicles move from closed race courses to city streets, they’ll have to judge the skills and intentions of other drivers, pedestrians and even the occasional squirrel in many situations.

Poker also is relevant to “games of national importance,” one of the topics that DARPA asked its Information Science and Technology Board to address. Some of those games have military significance—the feint by the Allies at a landing at Calais before launching their invasion of Normandy is a classic example—but others might address information warfare or international finance.

“So many interactions in the world are games—it just comes down to whether you can model them,” Sandholm says.

As in life, “luck plays a large role in poker,” Gilpin says. But it’s possible to minimize the role of chance when evaluating poker algorithms because the computer can remember each hand it dealt. In the AAAI competition, each two-player match consisted of 10,000 hands—5,000 as originally dealt, and 5,000 with the cards reversed, with pokerbot A getting pokerbot B’s original hands and vice versa.

It’s not as easy to filter luck out of a multiplayer game, Gilpin points out, because seating order is such a big factor. A player sitting to the left of a weak player, for instance, is at an advantage, while a player to the right is at a disadvantage. So, in the AAAI competition, seats were randomly

assigned prior to each of the twenty-five 6,000-hand matches.

One of the challenges of poker is simply whittling down the size of the game to something that is computationally manageable. Heads-Up Limit Texas Hold 'Em, a two-player game in which the size of each bet is restricted, has 10^{18} —or a billion times a billion—possible combinations of hands and bets. For no-limit games, in which the amount wagered is limited only by a player's bankroll, the number of possible combinations is a staggering 10^{71} .

As complex as the no-limit game might be, the multiplayer game is a far greater challenge. "I think it's one of the big open problems in multi-agent systems and artificial intelligence," says Sam Ganzfried, a computer science Ph.D. student who joined Sandholm and Gilpin in the poker group in 2007 and helped devise a six-player pokerbot for the 2008 competition.

Sandholm's group has developed a number of ways of automatically abstracting the games, such as recognizing strategically similar hands and grouping them together, to make the problem manageable. Gilpin and Sandholm have recently developed algorithms that take advantage of the greater computing power available through the Pittsburgh Supercomputing Center.

Early attempts at pokerbots, including pioneering work at the University of Alberta, Canada, often attempted to encode the expertise of human card pros. Another approach is to use game theory—the branch of applied mathematics that analyzes interactions between entities—to develop strategy based simply on the rules of the game. Simplified poker games have been tackled with game theory since the 1940s. The Alberta group began experimenting with game theory on full-scale two-person Texas Hold 'Em in 2002 and Sandholm opted to use game theory exclusively when he delved into computer poker research in 2004.

"I haven't read a single poker book in my life," says Sandholm, who makes no claim to being much of a card player. But he's convinced that game theory gave his pokerbots an edge: "Absent computational limitations, computers will play poker optimally."

Sandholm's group introduced the approach of using an automated abstraction algorithm to the game, followed by the use of a custom equilibrium-finding algorithm. Though the Carnegie Mellon programs didn't notch a win in the AAAI tournament until this year, they have always been strong contenders and all of the competitive pokerbots now employ the Sandholm group's approach.

Computers certainly play the game differently than most human players. Pokerbots are typically programmed to seek a Nash equilibrium—finding an optimal strategy that you wouldn't change even if you knew everything your opponent was trying to do. "But human players aren't optimizing their play," says Michael Bowling (CS'96, '99, '03), leader of Alberta's Computer Poker Research Group. "Their goal is to extract money from weaker players."

Drastic advances have been achieved in the past three years in both abstraction algorithms and equilibrium-finding algorithms, Sandholm says. For example, in 2006 the largest two-person zero-sum games that could be solved were those with no more than 10 million decision points; today, games 100,000 times bigger can be solved nearly optimally. And the supremacy of human players has begun to yield. Last summer, just prior to the AAAI conference, an Alberta pokerbot with the ability to adapt its play to its opponent's style edged out a couple of human poker champions in a human versus machine Heads-Up Limit Texas Hold 'Em competition in Vegas, just a year after narrowly losing to a couple of pros.

But Bowling says it's too early for pokerbots to declare victory over humans. "The fact is, we just don't know how strong human players are," he says. In repeated matches, humans might alter their style of play to exploit weaknesses in the pokerbots. And last year's win was in "limit" poker, not the more popular and more difficult "no-limit" game.

When it comes to no-limit, "we're good enough to play at the lowest professional ranks or at the upper amateur level," says Bowling, whose pokerbots won three out of the four categories in the 2008 AAAI contest, including the no-limit competition. Pokerbots for the limit game were operating at that level about two or three years ago.

Though Alberta has long dominated computer game research, the entry of Carnegie Mellon into the computer poker field since 2005 was something that the Alberta researchers welcomed because it helped sharpen the competition, Bowling says. "If you look at how much progress has been made on computer poker worldwide in the last three years, it's probably five times as much as the previous 12 years combined."

Which, as Sandholm sees it, means there's still a long way to go. "Humans are just one benchmark along the way," he says. "In 1997, a computer beat the world champion in chess, but that didn't mean that computer chess programs played chess optimally." To achieve the goal of optimal play, pokerbots will need improved abstraction algorithms so they can create a closer approximation of the real game, he says. Also, achieving a Nash equilibrium may not be sufficient; in multiplayer games in particular, it may prove impossible to reach a state where all players have optimized their strategy. As player modeling improves, computers may steal a trick from their human cousins and find that colluding against a weaker opponent may have advantages in some instances.

"We haven't been focusing on exploiting humans," Sandholm says. "But if some human wants to challenge us on a heads-up limit game, we'll put money on it."

—Veteran technology journalist Byron Spice is codirector of public relations for the School of Computer Science. Email him at bspice@cs.cmu.edu.



An Introduction to Data-Intensive, Scalable Computing

Our world is awash in data. Millions of devices generate digital data, an estimated 1 zettabyte (that's 10^{21} bytes) per year. Much of this data gets transmitted over networks and stored on disk drives. Thanks in part to dramatic cost reductions in magnetic storage technology, we can readily collect and store massive amounts of data. But what is all this data good for? Consider the following examples:

- At Wal-Mart, more than 6,000 stores worldwide record every purchase by every shopper, totaling around 267 million transactions per day. They collect this information in a 4-peta-byte (that's 4×10^{15} bytes) data warehouse set up for them by Hewlett-Packard. These transactions are a treasure trove of information about the shopping habits of their customers. How much did a \$200 discount on large-screen TVs increase sales, and how much did the shoppers who bought them spend on other things? How many copies of the upcoming John Grisham novel should they stock? Based on long-term weather forecasts, how many snow shovels should they order for their stores in Iowa? Sophisticated machine-learning algorithms can find answers to this question, given the right data and the right computing power.
- The proposed Large Synoptic Survey Telescope (LSST) will scan the sky from a mountaintop in Chile, with what can be considered the world's largest digital camera, generating a 3,200-megapixel image every 15 seconds, covering the total visible sky every three days. That will yield 30 terabytes (10^{12} bytes) of image data every day. Astronomers anticipate being able to learn much about the origins of the universe and the nature of dark matter by analyzing this data.
- Every time we have a CAT scan or an MRI taken, millions of bytes are recorded, but currently they're simply turned into a series of cross-sectional images. Imagine a future where

the different components of a knee joint could be identified, compared with the data from previous images, and a plan for knee surgery could be generated automatically.

These are just three of dozens of examples where an ability to collect, organize and analyze massive amounts of data could lead to breakthroughs in business, science and medicine. Of course, search engine companies have already demonstrated this capability for the information available on the World Wide Web, and they've shown they can make good money doing so. But the Web is just one of many possible data sources in our world, and search engines are just one form of data aggregation.

Storage Costs Almost Nothing; Analysis Remains Pricey

At one time, the cost of storing large amounts of data was prohibitive. Not any more. Modern disk drives have capacities measured in terabytes and cost less than \$100 per terabyte. A digitized version of all of the text in all of the books in the Library of Congress—essentially the totality of all of humankind's knowledge—would only constitute around 20 terabytes. Of course, data in the form of images, sound and video have much higher storage requirements, but still we can think of storage as being an almost-free resource. The challenge is in how to manage and make the most use of all of this data.

At Carnegie Mellon, we've made Data-Intensive, Scalable Computing (DISC) a major focus of our

research and educational efforts. We believe the potential applications for data-intensive computing are nearly limitless; that many challenging and exciting research problems arise when trying to scale up systems and computations to handle terabyte-scale datasets; and that we need to expose our students to the technologies that will help them cope with the data-intensive world they will live in. Realizing the promise of DISC requires combining the talents of people from across many disciplines, and Carnegie Mellon—with its strengths in engineering, computer science and many related fields—is uniquely qualified to lead the charge.

If we can fit the entire Library of Congress on 10 or 20 disk drives at a cost of \$2,000, what possible technical barriers can there be in realizing low-cost DISC systems and applying them to real-world problems? The core challenge is that communication and computation are much more difficult and expensive than storage. Consider that one-terabyte disk drive that we can buy for \$100. Even if we accept the optimistic claims of the drive manufacturer, we can only read or write 115 megabytes per second between that drive and a computer. That means it would take more than two hours just to scan the entire disk. Similarly, it would take almost the same amount of time to transfer that data between two computers connected by a high-speed, local-area network, assuming a transfer rate of one gigabit per second. Transferring a terabyte across a slower link, such as the Internet, could take many hours or even days.

So, if we had a terabyte dataset stored on a single disk, then answering even a simple query, such as the average of all of the values, would require several hours; applying more sophisticated analysis algorithms would be out of the question.

To deal with large-scale datasets, we need to spread the data across many disk drives, possibly hundreds, so that we can access large amounts of information in parallel. These disks also need processors and networking, and so we should incorporate

If we can fit the entire Library of Congress on 10 or 20 disk drives at a cost of \$2,000, what possible technical barriers can there be in realizing low-cost DISC systems and applying them to real-world problems?

them into a cluster computing system, comprising around one hundred nodes, each having one or two processors, several disk drives and a high-speed network interface. Creating this cluster requires construction of a machine room, with the nodes and power supplies mounted in racks and connected by cables. Our proposed \$2,000 storage system for the Library of Congress has now grown into a large-scale computing system costing nearly \$1 million.

Programming and operating systems with hundreds of disks and processors is no small task. One problem is reliability. A typical disk drive has a mean “time to failure” of around three years. For the disk drive in my laptop, every three years or so I have to deal with the inconvenience of having the drive replaced and restoring its contents from backup storage. (Hopefully!)

If we have a system with 150 disk drives, we can expect on average that one will fail every week. As we scale up a system, we must anticipate that any of its components—disks, processors, network connections, power supplies, even software—can fail at any time. Rather than stopping the system every time something goes wrong, we must engineer it to be highly fault tolerant, so that the system can continue operating (perhaps with some degradation in performance) despite multiple component failures. On the programming side, people have been trying for decades to use parallel computing to improve program performance, yet it remains high on the list of “grand challenge” problems for computer science. The combination of high performance, high reliability and ease of programming for parallel computing systems remains elusive.

Computing in the Cloud

Instead of building and operating our own clusters, one attractive approach is to make use of cloud computing systems. The idea is to let a dedicated organization take on the task of assembling a large system and making it available to others. For data-intensive computing, this can be in the form of a “virtual computing platform,” as exemplified by Amazon Web Services (AWS), where customers can buy computer cycles by the CPU-hour and network-accessible storage by the gigabyte-month. (The other cloud-computing model, referred to as “software as a service,” provides network-accessible applications, such as email or contact management. This is a valuable

Instead of building and operating our own clusters, one attractive approach is to make use of cloud computing systems. The idea is to let a dedicated organization take on the task of assembling a large system and making it available to others. For data-intensive computing, this can be in the form of a “virtual computing platform,” as exemplified by Amazon Web Services (AWS), where customers can buy computer cycles by the CPU-hour and network-accessible storage by the gigabyte-month.

service but not suitable for our needs.) From the clients’ perspective, cloud computing has the advantage that they do not need to worry about replacing disk drives, dealing with power outages or updating the operating system software on the nodes with the latest patches. Moreover, they can scale up their computing and storage capacity without building and provisioning new data centers.

On the software side, an open-source framework called “Hadoop,” styled after the system created by Google to run its Web crawling and indexing, makes it fairly easy to develop applications that manage and analyze large amounts of data on cluster systems. Yahoo! has been the major contributor to the development of the Hadoop File System. Hadoop handles the difficult issues of coordinating the activities of the node processors and disks to implement a large-scale, reliable file system. For example, most files are stored in triplicate on multiple disks, so that there are backup copies of every file if a disk drive fails. Programmers can write code that operates on data spread

across thousands of files via the MapReduce model pioneered by Google.

With this approach, the programmer describes computations to perform on the individual files (map), as well as how to combine the results of these computations (reduce) to produce data in the form of multiple files, with a typical program consisting of a sequence of MapReduce steps. The runtime system then handles the low-level details, such as scheduling the map and reduce tasks on the cluster processors and rerunning any tasks that fail. This frees the programmer from the traditional parallel programming concerns of data placement, scheduling and failure recovery. Software like Hadoop makes it possible to implement scalable and reliable applications running on otherwise unreliable and difficult to manage hardware.

Several companies have made their systems available for use by university researchers and educators. At Carnegie Mellon, we’ve been fortunate to have access to M45, a cluster system owned >>>

and operated by Yahoo!, to foster the growth of Hadoop and other open-source projects. (See “Mapping a New Paradigm,” *The Link*, Issue 3.1.) More recently, Google and IBM have joined forces to make large-scale machines available to U.S. universities under the auspices of the National Science Foundation. A number of universities are making use of the platform provided by Amazon Web Services, and Microsoft is developing systems and programming tools that are well suited for data-intensive computing.

New Models of Research, Training

Why are these companies providing access to their computer clusters? For one thing, universities supply students who will soon be working on these very systems. If students never see anything more complex than a program running on a single machine and datasets of one gigabyte or less,

Data-intensive, scalable computing is revolutionizing our ability to gather and analyze information in all forms. It will lead to new discoveries in science and new forms of entertainment. It will lead to improvements in business practices—and, just like any other disruptive technology, it will transform many different businesses for better and worse.

they won't be prepared to support and make use of large-scale, data-intensive computing. They won't have even thought about the insights that can be gained from analyzing terabytes of data.

There's a benefit to universities beyond education. We're interested in working on computing problems where the amount of data is so large that we can't own and operate the machines that would store and process it. For instance, a group of researchers in the Language Technologies Institute, led by Maxine Eskenazi, has been working to gather documents that can be used to teach English as a second language and in English education for elementary-school pupils. (See “Unraveling Language,” page 13.) Although there are plenty of documents available on the Web that could be used for reading practice or instruction, making

sure they're suitable for different reading levels isn't easy. Maxine and her team had to sift 200 million Web pages just to collect 200,000 useful documents. Only large-scale, data intensive computing makes that possible.

Another research group, led by Stephan Vogel of the Language Technologies Institute, is improving machine translation of one language to another. The modern approach to translating texts from, say, Chinese to English has been to scan millions or billions of documents in both languages, and then build statistical models that look for certain combinations of words and make their best guesses at what English words and phrases match those in Chinese. The success of these translation programs is greatly improved when they can be trained with more and more data—trillions of words. Without a large, scalable cluster of computers, dealing with that amount of data would be impossible.

To cite a third example, Alexei Efros of the Computer Science Department and the Robotics Institute, along with his Ph.D. student James Hay, has downloaded six million landscape images from Flickr and set up a program that looked for common features. Using geographic data attached to some of those photos, they then created computer models that were able to identify—with a surprising level of accuracy—where the different images were taken.

Beyond making use of existing clusters and software frameworks, there are many important research problems to be addressed to fully realize the potential of data-intensive computing. How can processor, storage and networking hardware be designed to improve performance, energy efficiency and reliability? How can we run a collection of data-intensive computations on the system simultaneously? What programming models and languages can we devise to support forms of computation that don't fit well in the MapReduce model? What machine-learning algorithms can scale to datasets with billions of elements? As a research organization, the School of Computer Science views DISC as a source of a large number of exciting opportunities.

There are advantages that universities bring to research in data-intensive computing that companies cannot match. First, the research we do is completely open. We publish our findings and share information freely, which private industry seldom does. Second, history has shown over and

over again that when you give creative people new capabilities, they will come up with new ideas that the originators never imagined were possible. The World Wide Web, after all, was never one of the original applications imagined for the Internet, but its impact has been transformative.

There are also research problems that can benefit society but have no prospect of forming a profitable business. Companies will not profit from basic scientific research in astronomy, for instance, but that research will enable us to deepen our understanding of the universe.

Looking Ahead to Broader Challenges

Within the next five years, more and more research projects at Carnegie Mellon will involve analyzing huge amounts of data on very large-scale machines. Our educational efforts are broadening as well. We've created courses at both the undergraduate and graduate level that give students a chance to run programs on the cluster provided by Google and IBM. We're also involved in several efforts to help the National Science Foundation provide course and teaching materials to get students at other universities involved in large-scale, data-intensive computing.

Data-intensive, scalable computing is revolutionizing our ability to gather and analyze information in all forms. It will lead to new discoveries in science and new forms of entertainment. It will lead to improvements in business practices—and, just like any other disruptive technology, it will transform many different businesses for better and worse. Data-intensive, scalable computing will change the way we think about science—the ability to collect and make use of so much data will change our perspective on the kinds of scientific and medical experiments that we consider attempting.

Data-intensive, scalable computing will change education and research in computer science at Carnegie Mellon profoundly. As our students go off into the world, they must understand this new paradigm in computing. They need to know how to use the technology and fully exploit the capabilities that large-scale computing make possible. At the School of Computer Science, we are committed to presenting our students with these new possibilities and providing them with the right kind of tools and thinking to help this technology advance.

Communication and Advocacy

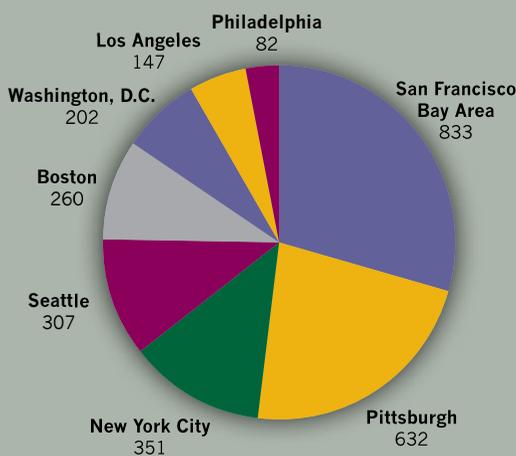
SCS ALUMNI AT-A-GLANCE

Total Alumni: 4,768
Male: 3,884
Female: 884

ALUMNI BY GEOGRAPHY

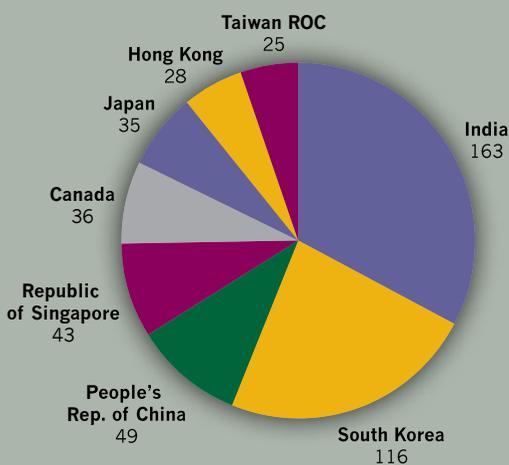
Domestic

Top Eight Locations



International*

Top Eight Locations



*Qatar class of 2008: 12



Tina M. Carr
 (HNZ'02)
 SCS Director of Alumni Relations
 tcarr@cs.cmu.edu

In this special “guest column,” SCS alumnus and Associate Dean Philip Lehman offers a few thoughts on the importance of communication and advocacy. —Tina

In the last issue of The Link, Alumni Relations Director Tina Carr discussed the “Gift of Time”—the gift given by alumni and friends of SCS who participate and volunteer. You attend events, give talks, speak with prospective students, participate in advisory boards and volunteer for many other initiatives.

These contributions help us continue to build success, and at the heart of these activities are communication and advocacy.

SCS communicates with you, our alumni and friends, to keep you informed about various aspects of campus life, our progress and achievements and our outreach and involvement with the world beyond campus. We use The Link, email, the Web, letters, the news media and direct, personal interaction. You provide your thoughts, suggestions and advice, and you visit with us, either on campus or at Carnegie Mellon events around the country. All of these are important, and we’re in the process right now of looking at the ways we communicate with you, and understanding where we can improve our communications strategies.

But there’s another type of communication that’s key to our mission and to getting

our message out. Every day, you talk with colleagues, co-workers and friends in industry, academia, government and the community. If you’re reading this column, it’s likely that some of your conversation is about Carnegie Mellon and SCS.

When you talk about us—with pride, we hope—you help us communicate, and you are advocating for us. The greater Carnegie Mellon and SCS community—including our alumni and friends—are our best advocates. Thank you for your advocacy. We encourage you to continue to be our strongest voice.

How can we do a better job of communicating with you? And how can we help you do a better job of discussing the innovative education and research underway at the School of Computer Science and Carnegie Mellon?

Please let us know your thoughts and ideas. We look forward to continuing the conversation with you!

—Philip L. Lehman (CS’78, ’84)
 Associate Dean for Strategic Initiatives
 School of Computer Science
 Philip.Lehman@cs.cmu.edu



Alumni Snapshots



Ashley Stroupe

Ashley W. Stroupe

B.S., Physics, Harvey Mudd College, 1990

M.S., Electrical and Computer Engineering, George Mason University, 1998

M.S., Robotics, Carnegie Mellon University, 2001

Ph.D., Robotics, Carnegie Mellon University, 2003

Ashley Stroupe was a small child when Apollo 17—the final manned mission to the moon—blasted off in 1972. But she can vividly remember watching the television coverage with her parents, including her father, an aeronautical engineer who worked on the space program. Stroupe decided she was going into space some day.

To her chagrin, she soon learned “they won’t let you be an astronaut if you get motion sickness on a carousel.”

Yet for the past five years, Stroupe has been driving on the surface of another planet. As a senior engineer at NASA’s Jet Propulsion Laboratory, she writes command sequences that control the

navigation and robotic arms of the two Mars rovers, Spirit and Opportunity, which were launched within weeks of each other in January 2004. Although the battered rovers are now five years past their original 90-day mission, they continue to send back valuable data about soil conditions on Mars.

Uploading instructions to the rovers takes only a few minutes, but learning whether they responded correctly takes up to 12 hours. “It can be stressful, especially when we’re trying to get out of an emergency or a trouble spot,” Stroupe says. “But the rovers haven’t disappointed us yet. The good nights far outweigh the bad.”

She credits current and past SCS faculty, including Manuela Veloso, Herbert A. Simon Professor

of Computer Science, with helping her chart a career path. “Manuela is a force of nature and infects everyone around her with this love for what she’s doing,” Stroupe says. Reid Simmons, research professor of robotics, was another inspiration, she says. And she calls former Robotics Institute research scientist Tucker Balch—now at Georgia Tech—her cheerleader: “He really encouraged me and helped me meet the right people.”

Though her eyes are on the stars, Stroupe keeps her feet on the ground. Away from JPL, she’s a mentor to middle school girls in Pasadena who are dealing with peer pressure and other problems of young adulthood.

—Jason Togyer (HS’96)

Tiffany Chang

B.S., Computer Science and Human-Computer Interaction, Carnegie Mellon University, 2003

M.B.A., Business Administration, Harvard Business School, 2008

No one knows the value of a good support network like Tiffany Chang. As an undergraduate majoring in computer science and human-computer interaction, she was one of the founding student members of Women@SCS, which promotes academic and professional interaction for women in a field once dominated by men.

The leadership and encouragement of faculty members Lenore Blum and Carol Frieze was crucial to the group’s early successes, Chang says. “They really made sure we had the resources we needed, and also helped spread the word.” One of the organization’s first outreach efforts sent SCS students into Pittsburgh area middle and high schools to encourage girls to consider careers in science and technology. Another matched female undergraduates with doctoral students to provide them with research mentors, Chang says.

“We wanted to make sure that women in the program felt comfortable asking questions,” she says. “We had to make sure they knew they were part of a community that was bigger than them, and make sure they had access to the help and material they needed.”

Chang’s personal support network has remained strong since graduation. While working on her M.B.A. at Harvard last year, she needed to draft a business plan and was allowed to choose two outside people to help. Chang picked classmates from Carnegie Mellon—Brandon Weber (HS’03) and George Davis (CS’03,’07), currently a doctoral student in the Institute for Software Research.

“The network you build at Carnegie Mellon is invaluable,” Chang says. “The people I met there were some of the smartest people I know.”



Tiffany Chang

Currently a manager for Palo Alto, Calif.-based VMware, a developer of virtual machine software, Chang says combining business school training with a computer science degree has provided her with a powerful career combination. “Technology now is so pervasive that literally anything you want to do would have some computer science in it,” she says. “I’ve never regretted doing it. It’s a great base for anything you want to do in the future.”

—JT

CS theorist Karp receives Dickson Prize

One of the leading theorists in computer science is this year's winner of Carnegie Mellon's Dickson Prize in Science. At a ceremony March 25 at the McConomy Auditorium, the university honored Richard M. Karp of the University of California at Berkeley. A lecture by Karp entitled "The Mysteries of Algorithms" followed.

A past winner of the A.M. Turing Award, Karp is a University Professor of Electrical Engineering and Computer Science at Berkeley and a research scientist at the International Computer Science Institute. His most recent research uses computational biology to understand how genes and living cells work.

Karp is renowned for his contributions to the Theory of NP-Completeness, a cornerstone of modern theoretical computer science that revolutionized algorithm design and paved the way for the integration of computing into scientific research.



Richard M. Karp

The \$50,000 Dickson Prize in Science is one of Carnegie Mellon's most prestigious honors. Created by the wills of Pittsburgh physician Joseph Z. Dickson and his wife, Agnes Fisher Dickson, it annually honors the American scientist judged

to have made the most progress in their field during the previous year.

The first Dickson Prize in Science was awarded in 1970. Through the same bequest, the University of Pittsburgh each year awards the Dickson Prize in Medicine.

Kiesler receives SIGCHI's highest honor

Sara Kiesler has been honored with a Lifetime Achievement Award from the Association for Computing Machinery's Special Interest Group in Computer-Human Interaction. Kiesler will receive a \$5,000 honorarium at SIGCHI's April meeting.

A member of Carnegie Mellon's faculty since 1979, Kiesler was named the university's Hillman Professor of Computer Science and Human-Computer Interaction in 2004. Her work includes research into the ways that groups of people form and interact over the Internet. Kiesler has studied such phenomena as "flame wars," distributed collaboration, information sharing and social equalization.

In 1998, Kiesler and Robert Kraut, Herbert A. Simon Professor of Human-Computer Interaction, reported that increased home access to email and the Web made users feel lonelier and less connected to their families. Their paper, Internet Paradox, received international media attention.

HCI doctoral student nets Smiley Award

A Web application called "Graffiter" that allows Twitter users to collect personal information and display it as a graph or chart is the winner of the 2009 Smiley Award.

Its creator, Ian Li, a doctoral student in the Human-Computer Interaction Institute, received the \$500 first prize Feb. 20. Sponsored by Yahoo! Inc., the prize recognizes communication innovations developed by Carnegie Mellon students. It's named for the ubiquitous smiley "emoticon" invented 26 years ago by Scott Fahlman, research professor of computer science.

Graffiter employs the "hash" tags built into Twitter to collect user-defined data about things such as hobbies, diets and exercise programs. "If there is something about your life that you're curious about, start recording it and study your graphs," Li said.

Fahlman, who presented Li with his award and a crystal trophy, called Graffiter "a viral application likely to spread quickly" through the Twitter community and "very much in the spirit" of the smiley emoticon.



GLENN BROOKES PHOTO

Scott Fahlman, research professor of computer science; doctoral student and Smiley Award winner Ian Li; and Don McGillen, senior campus relations manager for Yahoo! Inc., after the Feb. 20 ceremony in the Perlis Atrium of Newell-Simon Hall.

New Disney fellowship named for Pausch



Following the Feb. 4 dedication of a plaque honoring the late Randy Pausch (CS'88), his sons Logan and Dylan get ready to take cousins Micajah and Hannah Mason for a spin on the Mad Tea Party ride at Walt Disney World in Florida. Behind them, Pausch's wife, Jai, holds daughter Chloe as "Alice" and "The Mad Hatter" look on.

The Walt Disney Co. has created a fellowship in memory of the late Randy Pausch (CS'88), professor of computer science, human-computer interaction and design and one of the founders of the Entertainment Technology Center. Robert Iger, Disney president and chief executive officer, made the announcement Feb. 4 at Walt Disney World near Orlando, Fla.

The Disney Memorial Pausch Fellowship will support two graduate students—one each in the School of Computer Science and the College of Fine Arts—who are studying the interaction of art and technology.

Pausch, who died July 25, 2008 from pancreatic cancer, worked as a Walt Disney Imagineer in 1995 and later served as a consultant. He became an international celebrity after the "Last Lecture" he delivered at Carnegie Mellon in September 2007 was viewed online by hundreds of thousands of people and turned into a bestselling book.

Disney also unveiled a plaque honoring Pausch. Located next to the Mad Tea Party ride at Walt Disney World, it features a quote from the "Last Lecture": "Be good at something; it makes you valuable. Have something to bring to the table, because that will make you more welcome."

Degree trains Pittsburgh-area robotics technicians



Carnegie Mellon is leading a regional consortium in development of a new associate degree in robotics engineering technology. The two-year program will train technicians to build and maintain robots and other embedded computer systems, said Robin Shoop, director of Carnegie Mellon's Robotics Corridor and head of the Robotics Academy, an educational outreach effort inside the Robotics Institute.

Courses will be offered this fall at three Pittsburgh-area colleges, and credits will be transferable to four-year programs in electrical, computer and manufacturing engineering. Graduates will be qualified to work as managers or technologists in a variety of technology-intensive industries, Shoop said.

Funded by the National Science Foundation, the U.S. Department of Defense and other agencies, the Robotics Corridor is dedicated to creating the infrastructure necessary to support the robotics industry in the Pittsburgh region.

"SCS in the News" is compiled from staff reports.



FRC 25 + Red 60 =

Carnegie Mellon celebrated the 25th anniversary of the Field Robotics Center and the 60th birthday of its founder, William “Red” Whittaker, during the university’s Homecoming festivities.

On Friday, Oct. 24, Whittaker—Fredkin Professor of Robotics, director of the Field Robotics Center and founder of the National Robotics Engineering Consortium—delivered a free public lecture in Wean Hall entitled “Robots at Work.”

The Robotics Institute hosted some of the most distinguished experts in the field on Saturday, Oct. 25 for a day-long symposium. Speakers included David Akin of the University of Maryland; Hugh Durrant-Whyte of the University of Sydney; David Lavery of the National Aeronautics and Space Administration; Robin Murphy of Texas A&M University; Marc Raibert of Boston Dynamics; John Reid of Deere & Company; Sebastian Thrun of Stanford University; Brian Wilcox of Jet Propulsion Laboratory; and Dana Yoerger of Woods Hole Oceanographic Institution.

(Video and biographies of the speakers are available at www.fr25.org.)

A reception and banquet followed at Phipps Conservatory. At left, Whittaker autographs one of the table favors for Murphy. Both are wearing coonskin caps to celebrate Whittaker’s status as a robotics “pioneer.” Below, a shot of the symposium in the Giant Eagle Auditorium of Baker Hall.

SCREENSHOTS



4305 Newell-Simon Hall
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Then and Now

It's early Wednesday morning, March 28, 1979, in the small towns south of Harrisburg, Pa. Suddenly, sirens are piercing the quiet and firefighters are scrambling into action. At the Three Mile Island nuclear generating station, the Metropolitan Edison Company has declared a "general emergency."

The worst U.S. accident in the two-decade history of commercial nuclear power has come dangerously close to releasing a life-threatening cloud of radioactive material.

Five years later, the contamination in TMI's Unit 2 reactor was still too hazardous for humans to enter. At Carnegie-Mellon University, William "Red" Whittaker (E'75,'79), a native of central Pennsylvania, and graduate student Jim Osborn (E'81,'86) led a team that designed a robot to probe the basement of the reactor building.

Their six-wheeled Remote Reconnaissance Vehicle, dubbed "Rover," sent back images of the reactor's core, showing for the first time that it had suffered a partial meltdown. The Whittaker team's next robot, Remote Core

Sampler, brought back samples from the damaged walls; a third vehicle was designed to work inside the contaminated area.

The experience proved that robots had jobs outside of labs and factories and in the real world. From that realization, the Field Robotics Center was born.

Twenty-five years and many robots later, FRC's creations have probed the Antarctic, abandoned mines and the deepest sinkhole on the planet. They also proved the ability of autonomous vehicles to safely drive themselves on public streets when Boss won the DARPA Urban Challenge in 2007.

And if Whittaker has his way, a robot designed by FRC scientists will someday win the Google Lunar X Prize when it lands on the moon.

In October, faculty, students and roboticists from around the world gathered in Pittsburgh to celebrate the FRC's 25th anniversary and Whittaker's 60th birthday. Look inside for photos from the two-day event.



Rover and Red, 1983

FILE PHOTO