the newsletter of the International Computer Science Institute

The ICSI GAZETTE

volume nine | issue one | september 2010

featured icsi researchers:

richard karp

Richard Karp wears too many hats for one title to apply. He has appointments in UC Berkeley's departments of Electrical Engineering and Computer Science, Mathematics, Bioengineering, and Operations Research. He's the head and founder of ICSI's Algorithms Group (originally the Theory Group), and a member of the Networking Group. His interests span optimization, genomic modeling, and applied research.

"The problems almost always involve the design of combinatorial algorithms," Karp says.

Born in Boston, Massachusetts, the son of a junior high school math teacher, Karp was educated at the Boston Latin School and Harvard University. After receiving his Ph.D. in 1959, Karp began working at IBM in Yorktown Heights, New York, where, he says, "Most of the time I could just play around." At night, he taught evening courses at colleges around New York.

Karp then left IBM to become a professor at UC Berkeley. "This was during the 1960s, when it seemed more fun to be in Berkeley than anywhere else," he said. "It was a very pleasant culture shock." He remembers colleagues being jailed for participation in the protests, and having to hold classes in his home when political protests shut the campus down

He has been with ICSI since 1988, when the Institute opened for business. As the Algorithms Group leader, he supervises researchers from all over the world interested in a wide cross-section of topics related to theoretical algorithms. In the 2009 to 2010 academic year, nearly half of his group were postdoctoral Fellows sponsored by the German Academic Exchange Service. He also collaborates closely with researchers in and from Israel, which, he says, accounts for 20 percent of the world's activity in theoretical computer science.

continued on page 4

charles fillmore

Charles Fillmore, director of ICSI's FrameNet Project and professor emeritus of linguistics at UC Berkeley, has studied language in wildly different settings over the course of his sixty-year career.

Fillmore received his undergraduate degree at the University of Minnesota, after which he spent three years in the U.S. Army stationed in Japan, intercepting coded Russian conversations on short-wave radio. When he was off duty, he walked around the streets with a notebook and worked at teaching himself Japanese. After he was discharged – the first U.S. soldier to be discharged locally in Japan – he taught English at a Buddhist girls' school while taking classes at Kyoto University. He returned to the U.S. to receive his doctorate at the University of Michigan, and then spent ten years teaching at Ohio State University in Columbus. He spent one year as a fellow at the Center for Advanced Study in Behavioral Sciences at Stanford University in what he calls "senseless luxury," and came to UC Berkeley in 1971.

According to FrameNet PI Collin Baker, Fillmore "has been a quiet voice of reason in the linguistics wars over the decades, because he's always been true to the data." While many linguistics theories attempt to describe all aspects of a language – working from the top down, as it were – Fillmore works from the data within a loose conceptual framework, allowing for exceptions to his theories. This flexibility forms the basis of Fillmore's brand of linguistics.

Fillmore has helped develop some of the fundamental concepts in linguistics as it is practiced today. In the 1960s and 1970s, he developed the theories of case grammar and frame semantics, and in the 1980s and 1990s he developed and contributed to work in construction grammar.

continued on page 5

in this issue:

as i see it by Nelson Morgan, Director

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.

- J.R.R. Tolkien

Recently I have become interested in the perception of science and engineering in our field. Consider the term "Computer Science"; hopefully we use scientific methodology in our field, but clearly much of what we do is engineering. Is this a "lower" pursuit? At Berkeley, CS is part of the EECS Department, which in turn is part of the College of Engineering. Are the CS faculty scientists, or engineers? Dictionary definitions of these terms are all over the map – including one that defines an engineer as a person who designs or builds engines – but it seems to me that these terms are best understood in the context of the goals of the professions. In practice, engineering focuses on making something work (or work better); and gaining knowledge (often about the natural world) is the fundamental goal of science. And in computer science, as well as in many other pursuits, the two often go together – as noted in the Tolkien quote above, we often start off with a goal of improving some functionality (like speech or object recognition) and in the process learn something quite different, such as a new approach to machine learning.

Scientific exploration often requires significant efforts in engineering; and good engineering practice incorporates the scientific method, and implicitly or explicitly takes advantage of scientific knowledge. Perhaps these are obvious points; yet it is common in our field to view the two disciplines as disjoint, with engineering being viewed as the lesser activity.

To be concrete, let's take the study of automatic speech recognition (ASR), a topic that I'm pretty well acquainted with. Our research goal in this area is to significantly improve the accuracy of ASR, particularly in conditions that break current systems, e.g., for noisy or reverberant acoustic environments. This is an engineering goal, which also typically has constraints that require tradeoffs, e.g., between computational requirements and accuracy. However, it has frequently been true that advances in ASR have required the development of new mathematical formulations or models, or have been inspired by an understanding of the perception or production of speech. Science and engineering have often been good partners in this area.

On the other hand, many of the advances in ASR have required significant efforts in the processing of large files, multiple types of data sets, attention to formats, etc., and this has often been viewed as "mere engineering." This activity can be critical, but it does make sense to think of it as being fundamentally different from the design of a new algorithm or model. Yet both can be viewed as engineering if the goal is to design or improve some system.

My point here is that much of engineering has science at its core, and much scientific inquiry can require significant engineering of a nontrivial variety. As Adam Janin of ICSI recently remarked to me, perhaps the more useful division is between research and development. The former implies the process of discovery, whether it concerns an artificial system or a natural phenomenon; while the latter suggests the incorporation of existing methods in order to complete the creation of some working entity. Faculty and staff at ICSI and places like it (although, really, there's no place like ICSI) are focused on research, though development is typically a necessary part of the effort; but some of this research is certainly best described as engineering research.

It may be that these naming conventions are of little import. But it is astonishing to me how often I see work that is quite original in nature denigrated as "mere engineering" while other work that is derivative is elevated as "science" simply because it has no practical application. Research is not always science, and engineering is not always development. Research at ICSI is both of the engineering and scientific variety, but in both cases what gives us our reputation is the outstanding standard of excellence.

Speaking of excellence, in this issue of the Gazette, we are focusing on two of the key members of the ICSI family, both eminent scholars in their respective fields: Charles Fillmore and Richard Karp. We're extremely fortunate to have them in our midst.

news briefs

ICSI Director **Nelson Morgan** has been named as a Fellow by the International Speech Communication Association. Morgan is honored for his significant contributions to robust feature extraction and novel acoustic models for automatic speech recognition. He is one of 24 Fellows named since the Fellow Program began in 2007, and the second ICSI Speech Group member to be recognized with the distinction – Elizabeth Shriberg was named in 2009. The six new Fellows were announced at this year's Interspeech Conference, held in September.

Orpheus Crutchfield, executive director of the **Berkeley Foundation for Opportunities in Information Technology** (**BFOIT**), was profiled in the August 1 edition of the *San Francisco Chronicle Magazine*. BFOIT is an ICSI program that supports historically underrepresented ethnic minorities and women in technology. Through seminars hosted at UC Berkeley and ICSI, BFOIT inspires local high school and middle school students to pursue careers in computer science, mathematics, engineering, and information technology. The Chronicle profile, "Orpheus Crutchfield is Lifelong Mentor," highlighted Crutchfield's commitment to promoting diversity and encouraging youth.

Algorithms Group graduate student **Bonnie Kirkpatrick** has won a 2010 Google Anita Borg Memorial Scholarship. The award, given annually to female students in technology and computer science for academic excellence and leadership, includes a $10,000 scholarship and invitations to networking retreats at Google. The award was established in 2004 to honor Dr. Anita Borg, the founder of the Institute for Women and Technology, and is dedicated to encouraging female scholars to become leaders in the fields of technology and computing.

Paul Kay, a researcher with the AI Group, co-authored an article in the June 1 issue of the *Proceedings of the National Academy of Sciences*. The article is Kay's tenth on color naming and perception to appear in PNAS in the last five years.

Vern Paxson, head of the security branch of the Networking Group, was quoted in an article on Internet security in the July/August edition of *MIT Technology Review*. The article, "Moore's Outlaws" by David Talbot, describes recent attacks by botnets, groups of computers that are infected with viruses. These botnets are controlled remotely and are responsible for the vast majority of spam on the Internet. Networking's research on botnets has also been featured in *New Scientist* magazine.

ICSI research in speaker diarization was presented at the Research@Intel 2010 Day on June 30. Speech Group researchers **Adam Janin**, **Gerald Friedland**, and group leader **Nelson Morgan** presented the Meeting Diarist, a tool that recognizes who is speaking and what is being said, and allows users to search for relevant parts of a meeting's transcript. *PC Mag* featured the presentation in a July 1 article about the event.

A UC Berkeley team led by Professor Pieter Abbeel, ICSI Vision Group leader **Trevor Darrell**, and Professor Stuart Russell is one of the winners in the Personal Robot 2 Beta Program, sponsored by robotics company Willow Garage. The team, which includes UC Berkeley Professor **Dan Klein**, a Speech Group affiliate, will be given a robot for two years, and will experiment with hierarchical planning, perception, and manipulation of deformable objects, such as towels.

Jörg Lässig and **Dirk Sudholt**, DAAD-sponsored postdoctoral fellows in the Algorithms Group, received two best paper awards this summer. At the International Conference on Parallel Problem Solving from Nature (PPSN) in September, the scholars won the award for their paper "General Scheme for Analyzing Running Times of Parallel Evolutionary Algorithms." At the Genetic and Evolutionary Computation Conference (GECCO) in July, their paper titled "The Benefit of Migration in Parallel Evolutionary Algorithms" was the best paper in the Parallel and Evolutionary Systems track.

Architecture Group leader **Krste Asanovic** has been given the 2009-2010 Jim and Donna Gray Faculty Award. The award is given each year to a UC Berkeley Computer Science faculty member for excellence in undergraduate teaching. Previous winners include ICSI affiliate Dan Klein (2008-2009) and BFOIT advisory board member Daniel Garcia (2004-2005).

Professor **Dan Klein** has been awarded a 2010 Distinguished Teaching Award, UC Berkeley's most prestigious award for instruction. The awards are given annually to recognize sustained excellence in the classroom, creative scholarly work, and the ability to inspire independent thinking in students.

New Scientist magazine and Heise's online security journal are encouraging their readers to use the Netalyzr system, developed by members of the Networking Group. **Netalyzr**, which has also been featured in *The Register*, analyzes the extent to which a user's Internet service provider interferes with its customers' traffic. To date, more than 130,000 users from 180 countries have tried the system. *New Scientist* also provides background information written for a nontechnical audience on the tests conducted by Netalyzr. The system was developed by **Christian Kreibich**, **Vern Paxson**, and **Nicholas Weaver** of the Networking Group.

Networking Group scientist **Christian Kreibich** was quoted by BBC News discussing the difficulties of filtering spam email. The article focused on harmless emails incorrectly marked as spam. Researchers at ICSI and UC San Diego recently developed a spam blocking method that, in initial tests, only filtered spam messages, with no false positives – that is, legitimate messages marked as spam. The system exploits the spam-creating botnets that run in the background of users' computers by learning the templates the bots use and teaching spam filters to look for these templates. While this doesn't guarantee a 100 percent success rate in real life situations, it is a huge step forward in spam prevention.

featured researcher: richard karp

continued from page 1

"Enjoy your failures as much as your successes, because they will be much more numerous."

Besides four years at the University of Washington, Karp has been at UC Berkeley since 1968, where he has earned a reputation as an excellent teacher. He was awarded a Distinguished Teaching Award from UC Berkeley.

"Teaching is first of all a lot of fun," he says. "It makes you rethink your work, and when you have to give a lecture on an idea, it forces you to understand it."

Even after decades of research, Karp is continually refining and rethinking his ideas. He's always been interested in puzzles and random choices, but when he took an undergraduate course at Harvard University in probability theory, he turned his attention in that direction. While a graduate student, he studied computational methods for solving optimization problems in industry. He's particularly interested in discrete mathematics. His research at IBM was mainly in algorithmic linear programming, in which the mathematician tries to find the best outcome for a problem given a series of constraints.

He eventually began to study computational complexity theory, and in 1972 published the landmark paper "Reducibility Among Combinatorial Problems." In that paper, he showed that some algorithms problems can be impossibly difficult to solve. He's interested in "why we can solve these problems even though complexity theory tells us it's impossibly hard." In awarding him the Turing Award in 1985, the ACM cited Karp for introducing the "now standard methodology for proving problems to be computationally difficult."

In the early 1990s, with the advent of the Human Genome Project, Karp became interested in applying combinatorial and probabilistic algorithms to biological problems. Much of the work today of ICSI's Algorithms Group focuses on answers to biological questions.

In recent years, he has returned to heuristic algorithms, algorithms which make sense intuitively but not theoretically, and problems related to machine learning. He has also become more and more interested in specific problems, rather than theories as a whole – for example, in looking at actual computer runs.

Over the years, his intellectual flexibility and groundbreaking research have brought him numerous accolades – including, notable, the Turing Award and the Kyoto Prize, which is described as the Japanese Nobel Prize. But for Karp, research is as much a process of failing as it is of succeeding – of trying different approaches over and over again without any guarantee of success.

He says, "Enjoy your failures as much as your successes, because they will be much more numerous."

featured researcher: charles fillmore

continued from page 1

"I've always been interested in meaning as much as form, and the relationship between form and meaning," he says.

During the early 1990s, he taught summer school classes in computational lexicography at the University of Pisa. There, he met Sue Atkins, a lexicographer with whom he continues to collaborate. At the time, Atkins was involved in a European Union-funded project that was using frame semantics to study perception verbs in English, French, Danish, Italian, and Dutch, and he joined the group as an external consultant. "I gradually convinced myself that we had to do something like that in Berkeley," he said.

Fillmore retired from UC Berkeley in 1995, and shortly afterwards, then-ICSI Director Jerome Feldman invited him to join the Institute and write a proposal for lexical semantic research. In 1997, the group received its first grant and began to build the FrameNet database, which shows English words in each of their senses and lays out the ways in which they combine with other words to build complete phrases. The FrameNet Project aims to cover as much as possible of what is involved in understanding text. Words are grouped according to the semantic frames – schematic representations of situation types (like eating, removing, etc.) – that they participate in, and the patterns in which they combine with other words and phrases near them are described according to how the frame elements get expressed. Frame elements are the things that are worth talking about when a frame has been activated by a word. For example, verbs of buying and selling need to connect with a buyer, a seller, some goods, and some money, either explicitly in the sentence or implicitly in the situation in which they are used; verbs of revenge have to involve an avenger, an offender, an injured party, and a punishment.

"The happiest time of my career has been here at ICSI," says Fillmore, "where FrameNet has made it possible for me to work with a team of bright young people on a continuing basis doing work that I'll never lose interest in."

Over a decade after its inception, the FrameNet Project continues to add to its database. The project has manually annotated over 150,000 sentences, and has defined more than 10,000 lexical units (a pairing of a word with a meaning), 6,000 of which have been fully annotated. The FrameNet Web site includes a searchable database that gives users lists of frames, frame elements, and example sentences for all of the possible meanings for each word. The team has also been adding descriptions of the formal and semantic properties of special grammatical constructions that have meanings that can't be predicted from the meanings of their parts. "By finding representative examples of the uses of each word and classifying the meanings of the phrases that go with them, we are able to include in our database information about the words in our language that standard dictionaries simply don't have room for," he says.

We're building information about English for both people and computers – for people who want information about English and the way the mind works. And people who want to design programs need to have this information available."

The FrameNet resource has been used in the creation of lexicons for other languages and in the development of various natural language processing applications such as question answering and machine translation. Other researchers have begun creating FrameNets for several languages besides English, including Chinese, Japanese, German, Spanish, and Hebrew. As they build these FrameNets, they are finding that different languages have many of the same frames and use words in semantically similar ways. The hope is that frame semantics will one day guide automatic machine translation.

The FrameNet database is in use by hundreds of researchers around the world, and Fillmore's work is continually being recognized by his colleagues. In 1991, he served as president of the Linguistic Society of America. In 2000, he was awarded an honorary doctorate from the University of Chicago; in August 2009, a three-day conference was held at UC Berkeley in celebration of his 80th birthday, with papers on frame semantics and construction grammar by dozens of his former students and colleagues. He was the subject of an interview published in this year's Review of Cognitive Linguistics, and in December 2009, he conducted the FrameNet Masterclass and Workshop at the Eighth International Workshop on Treebanks and Linguistic Theories in Milan, Italy.

featured research: geo-tagging

Photos and videos posted on Web sites such as Craigslist, Twitter, and YouTube can carry detailed information about where the images were taken. According to Speech Group researcher Gerald Friedland and Networking researcher Robin Sommer, this may leave those who post the images vulnerable to "cybercasing," the use of geo-tagged information available online to mount attacks in the real world.

While geo-tags – information about where a photo or video was taken – have been in use for years, the increasing amount of data available on the Internet, combined with easy-to-use search programs now offered by Web sites such as YouTube, make geotags an emerging threat.

It doesn't help, say the researchers, that people are unaware that this information is easy to find and that the geo-tags are extremely accurate.

Geo-tags are automatically embedded in images by higher-end digital cameras and smart phones, such as the iPhone and the Android. Friedland and Sommer cross-referenced the latitude and longitude contained in images' files with publicly available information, such as Google Maps Street View, to quickly find street addresses.

The researchers estimate that 4.3 percent of Flickr photos and 3 percent of YouTube videos are geo-tagged. This means that approximately 180 million Flickr photos and 3 million YouTube videos are tagged with longitude and latitude coordinates.

In one case study, the researchers monitored the Bay Area's For Sale section on Craigslist, where many users choose to hide their real names and email addresses. In four days, they collected over 900 images that were tagged with GPS coordinates.

To verify the accuracy of coordinates embedded in these photos, they took a picture of a bike against a garage with an iPhone 3G camera as though they were going to sell it on Craigslist. Putting the photo's location metadata into Google Maps Street View, the researchers were able to pinpoint, within one meter, the actual location of the bike.

Some Web sites allow users to search images' metadata through publicly available application programming interfaces (APIs). In one search of YouTube's API, the researchers looked for homes near downtown Berkeley by searching embedded geo-location data and including the search term "kids," since many home videos are of users' children. They then searched for videos posted by the same users that had been filmed over 1000 miles away. Within fifteen minutes and using a simple 240-line Python script, the researchers were able to find a resident of Albany, California who was vacationing in the Carribbean, along with a dozen other users who might be vulnerable to burglary given the subject matter of the videos they had posted.

The researchers warn that if technology such as text and image recognition continues to be improved, such processes could be automated.

The results were presented at the USENIX Workshop on Hot Topics in Security in August, and also featured by ABC's Good Morning America, *New York Times*, and *New Scientist* magazine.

An example of pictures that can be located using geotagged information, overlaid on a Berkeley map.

visiting scholars

Since its inception, ICSI has had a strong international program consisting primarily of ties with specific countries. Current formal agreements exist with Brazil, Finland, Germany, and Switzerland. In addition, we often have visitors associated with specific research and projects.

AI
Terry Regier
Bracha Rubin
Hiroaki Sato
Carlos Subirats (Spain)
Kyoko Ohara

Algorithms
Christian Hochmuth (Germany)
Oliver Kramer (Germany)
Jörg Lässig (Germany)
Shuai Cheng Li
Benjamin Satzger (Germany)
Dirk Sudholt (Germany)

Architecture
Sascha Hunold (Germany)

Speech
Joaquin Gonzalez

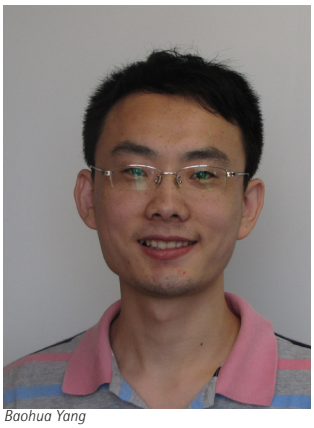
Networking
Kevin Bauer
Maria Bermudez-Edo
Arka Bhattacharya
Rainer Böhme (Germany)
Joos-Hendrik Böse (Germany)
Mobin Javed
Dmitriy Kuptsov
Katrina LaCurts
Matti Mantere (Finland)
Ha Young Oh
Jarno Rajahalme (Finland)
Pasi Sarolahti (Finland)
Petri Savolainen (Finland)

Campus Affiliation / Other
Dennis Brandão (Brazil)
Christoph Göbel (Germany)
Kai Huotari (Finland)
Tommi Lampikoski (Finland)
Jouni Similä (Finland)
Marko Turpeinen (Finland)

Vision
Nicolas Cebron (Germany)
Carl Henrik Ek
Sanja Fidler
Yang Hua (Panasonic/Singapore)
Zhong Yang Huang (Panasonic/Singapore)
Tinne Tuytelaars


Joaquin Gonzalez


Joos-Hendrik Böse


Baohua Yang


Kai Huotari

featured alum: matthew aylett

When Speech Group alum Matthew Aylett's company developed text-to-speech software that would allow Roger Ebert, the popular film critic who lost his voice in 2006 in his battle with thyroid cancer, to speak in an approximation of his own voice again, he expected a little media attention.

He didn't expect to be giving interviews to CBS, the BBC, and Reuters news service. That was the same week Ebert was on the front page of *Esquire* magazine and on the Oprah Winfrey Show, demonstrating his new voice for the first time.

Aylett received a bachelor's degree in artificial intelligence from the University of Sussex. After a five-year break, he returned to school to pursue his MSc and PhD at the University of Edinburgh. In 2000 he joined Rhetorical Systems, an Edinburgh company that produced engaging and colorful synthetic voices.

In 2004, after leaving Rhetorical Systems, he was sponsored for a visit to ICSI's Speech Group by the European Union through the Augmented Multi-Party Interaction (AMI) Project. "It was an important time," he says, "when I was asking what I was going to do next." ICSI, Aylett says, allowed him to return to an academic research setting while developing contacts with start-ups around the Bay Area. The environment at ICSI was "really conducive to creating stuff and getting stuff done."

He returned to Edinburgh in 2005 and helped found CereProc, which, like Rhetorical Systems, produces synthetic voices that retain local accents, have character, and are pleasant to interact with. While most companies are interested in synthetic voices being intelligible, says Aylett, "we're interested in it sounding good."

CereProc produces voices using unit selection. The company will record a voice actor reading a script. The recorded voice is transcribed and broken down into individual phonemes, which are reassembled when a user types a sentence into a program, producing spoken text. In Ebert's case, the company used audio from Ebert's DVD commentary.

Often speech synthesis companies will record speakers reading in a monotone voice to make transcription easier and the synthetic voice more intelligible. But, Aylett says, such voices have limited use as they are not engaging. "We take more risks with what we're doing," says Aylett. CereProc looks for interesting accents, from Ireland, Scotland, Southern England, America, and the Black Country in central England. In addition, CereProc is developing voices in French, Spanish, Catalan, German, Mandarin, and Japanese.

CereProc is also able to reproduce emotion by asking speakers to read scripts in a calm or tense voice. Aylett hopes, in future, to develop technology that allows emotion to be added to already recorded voices by changing the pitch and speech of synthesized voices.

Recently, CereProc has begun to produce voices using the HTS system, which uses a Hidden Markov Model to train software off of a small amount of recorded sound – minutes instead of hours. With the system, errors in transcription are less of an issue since the system trains itself, and producing an intelligible voice is less expensive, which means more people may be able to use CereProc's services. However, the system isn't perfect, producing voices with less character and softer accents than the unit selection system.

As technology becomes more pervasive and mobile, Aylett says, engaging synthetic voices will become more important. With the popularity of mobile devices like the iPhone, spoken information can be easier to absorb than written text.

Aylett also hopes in the future to help others who have lost their ability to speak, as they did with Ebert. "Our voice is profoundly part of who we are," he says.

publications listing

Z. Al-Qudah, M. Rabinovich, and M. Allman. Web Timeouts and Their Implications. Proceedings of the 11th Passive and Active Measurement Conference (PAM 2010), Zurich, Switzerland, April 2010.

M. Allman, K. Avrachenkov, U. Ayesta, J. Blanton, and P. Hurtig. RFC 5827: Early Retransmit for TCP and SCTP. Request for Comments 5827, Experimental, April 2010.

K. Asanovic and R. Wittig. Guest Editors' Introduction: Proceedings of the 21st Symposium on High Performance Chips (Hot Chips 21), Stanford, California. IEEE Micro, Vol. 30, No. 2, pp. 5-6, March 2010.

F. Bagci, J. Wolf, B. Satzger, and T. Ungerer. UbiMASS - Ubiquitous Mobile Agent System for Wireless Sensor Networks. Proceedings of the Third IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC2010), Newport Beach, California, pp. 245-252, June 2010.

M. Bansal and D. Klein. Simple, Accurate Parsing with an All-Fragments Grammar. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), Uppsala, Sweden, pp. 1098-1107, July 2010.

S. Beamer, C. Sun, Y.-J. Kwon, A. Joshi, C. Batten, V. Stojanovic, and K. Asanovic. Re-Architecting DRAM Memory Systems with Monolithically Integrated Silicon Photonics. Proceedings of the 37th International Symposium on Computer Architecture (ISCA 2010), Saint-Malo, France, pp. 129-140, June 2010.

P. Berenbrink, C. Cooper, R. Elsaesser, T. Radzik, and T. Sauerwald. Speeding Up Random Walks with Neighborhood Exploration. Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA 2010), Austin, Texas, pp. 1422-1435, January 2010.

R. Böhme and S. Köpsell. Trained to Accept? A Field Experiment on Consent Dialogs. Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI 2010), Atlanta, Georgia, pp. 2403-2406, April 2010.

R. Böhme and G. Schwartz. Modeling Cyber-Insurance: Towards a Unifying Framework. Proceedings of the 9th Workshop on the Economics of Information Security (WEIS 2010), Cambridge, Massachusetts, June 2010.

J.-H. Böse, A. Andrzejak, and M. Hoegqvist. Beyond Online Aggregation: Parallel and Incremental Data Mining with Online Map-Reduce. Proceedings of the Workshop on Massive Data Analytics on the Cloud (MDAC 2010), Raleigh, North Carolina, Article No. 3, April 2010.

K. Bringmann and T. Friedrich. Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects. Computational Geometry, Vol. 43, Issues 6-7, pp. 601-610, August 2010.

T. Callahan, M. Allman, and V. Paxson. A Longitudinal View of HTTP Traffic. Proceedings of the 11th Passive and Active Measurement Conference (PAM 2010), Zurich, Switzerland, April 2010.

N. Cebron. Refinement of Observations Under Budget Constraints in Active Learning. Proceedings of the Towards Closing the Loop: Active Learning for Robotics Workshop at the 2010 Robotics: Science and Systems Conference (RSS 2010), Zaragoza, Spain, June 2010.

A. Celikyilmaz and D. Hakkani-Tür. A Graph-Based Semi-Supervised Learning for Question Semantic Labeling. Proceedings of the Workshop on Semantic Search at the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference (NAACL HLT 2010), Los Angeles, California, pp. 27-35, June 2010.

A. Celikyilmaz and D. Hakkani-Tür. A Hybrid Hierarchical Model for Multi-Document Summarization. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), Uppsala, Sweden, pp. 1149-1154, July 2010.

A. Celikyilmaz, D. Hakkani-Tür, and G. Tur. LDA Based Similarity Modeling for Question Answering. Proceedings of the Workshop on Semantic Search at the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference (NAACL HLT 2010), Los Angeles, California, pp. 1-9, June 2010.

C. Y. Cho, J. Caballero, C. Grier, V. Paxson, and D. Song. Insights from the Inside: A View of Botnet Management from Infiltration. Proceedings of the Third USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '10), San Jose, California, April 2010.

J. Chong, G. Friedland, A. Janin, and N. Morgan. Opportunities and Challenges of Parallelizing Speech Recognition. Proceedings of the Second USENIX Workshop on Hot Topics in Parallelism (HotPar '10), Berkeley, California, June 2010.

J. A. Colmenares, S. Bird, H. Cook, P. Pearce, D. Zhu, J. Shalf, K. Asanovic, and J. Kubiatowicz. Resource Management in the Tessellation Manycore OS. Proceedings of the Second USENIX Workshop on Hot Topics in Parallelism (HotPar'10), Berkeley, California, June 2010.

D. Emig, N. Salomonis, J. Baumbach, T. Lengauer, B. R. Conklin, and M. Albrecht. AltAnalyze and DomainGraph: Analyzing and Visualizing Exon Expression Data. Nucleic Acids Research, Vol. 38, Web Server Issue, pp. W755-W762, July 2010.

J. Feldman. Cognitive Science Should Be Unified: Comment on Griffiths et al. and McClelland et al. Trends in Cognitive Science, Vol. 14, Issue 8, pp. 339-388, August 2010.

M. Felegyhazi, C. Kreibich, and V. Paxson. On the Potential of Proactive Domain Blacklisting. Proceedings of the Third USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '10), San Jose, California, April 2010.

M. Ferber, S. Hunold, and T. Rauber. BPEL Remote Objects: Integrating BPEL Processes into Object-Oriented Applications. Proceedings of the 7th International Conference on Services Computing (IEEE SCC 2010), Miami, Florida, July 2010.

O. Ferrández, M. Ellsworth, R. Muñoz, and C. F. Baker. Aligning FrameNet and WordNet Based on Semantic Neighborhoods. Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC 2010), Valletta, Malta, May 2010.

G. Friedland. Review of J. Nichols and B. Myers, "Creating a Lightweight User Interface Description Language: An Overview and Analysis of the Personal Universal Controller Project." ACM Computing Reviews, CR 137773, March 2010.

publications, cont.

G. Friedland. Review of E. Aguilar, "Animation and Performance Capture Using Digitized Models." ACM Computing Reviews, CR138181, July 2010.

G. Friedland and D. van Leeuwen. Speaker Recognition and Diarization. In Semantic Computing, P. Sheu, H. Yu, C. V. Ramamamoorthy, A. K. Joshi, and L. A. Zadeh, eds., pp. 115-130, IEEE Press/Wiley, 2010.

G. Friedland and R. Sommer. Cybercasing the Joint: On the Privacy Implications of Geo-Tagging. Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C., August 2010. Also appeared as ICSI Technical Report TR-10-005, May 3, 2010.

G. Friedland, O. Vinyals, and T. Darrell. Multimodal Location Estimation. ICSI Technical Report TR-10-007, May 8, 2010.

T. Friedrich and F. Neumann. When to Use Bit-Wise Neutrality. Natural Computing, Vol. 9, No. 1, pp. 283-294, March 2010.

M. Fritz, M. Andriluka, S. Fidler, M. Stark, A. Leonardis, and B. Schiele. Categorial Perception. In Cognitive Systems, H. I. Christensen, G.-J. Kruijff, A. Sloman, and J. Wyatt, eds., Springer, pp. 131-164, 2010.

M. Fritz, G.-J. Kruijff, and B. Schiele. Tutor-Based Learning of Visual Categories Using Different Levels of Supervision. Computer Vision and Image Understanding, Vol. 114, Issue 5, pp. 564-573, May 2010.

M. Gairing, T. Lücking, M. Mavronicolas, and B. Monien. Computing Nash Equilibria for Scheduling on Restricted Parallel Links. Theory of Computing Systems, Vol. 47, No. 2, pp. 405-432, August 2010.

P. B. Godfrey, M. Schapira, A. Zohar, and S. Shenker. Incentive Compatibility and Dynamics of Congestion Control. ACM SIGMETRICS Performance Evaluation Review, Vol. 36, Issue 1, pp. 95-106, June 2010.

U. Guz, G. Tur, D. Hakkani-Tür, and S. Cuendet. Cascaded Model Adaptation for Dialog Act Segmentation and Tagging. Journal of Computer Speech and Language, Vol. 24, Issue 2, pp. 289-306, April 2010.

E. Halperin. Deciphering the Genetic Components of Human Diseases. Invited talk at the 7th International Conference on Algorithms and Complexity (CIAC 2010), Rome, Italy, May 2010.

C. A. Hochmuth, J. Lässig, and S. Thiem. Simulation-Based Evolutionary Optimization of Complex Multi-Location Inventory Models. Proceedings of the Third IEEE International Conference on Computer Science and Information Technology (IEEE ICCSIT 2010), Chengdu, China, July 2010.

C. Horoba and D. Sudholt. Ant Colony Optimization for Stochastic Shortest Path Problems. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2010), Portland, Oregon, pp. 1465-1472, July 2010.

S. Hunold. Low-Cost Tuning of Two-Step Algorithms for Scheduling Mixed-Parallel Applications onto Homogeneous Clusters. Proceedings of the 10th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid 2010), Melbourne, Australia, pp. 253-261, May 2010.

A. Kapoor, K. Graumann, R. Urtasun, and T. Darrell. Gaussian Processes for Object Categorization. International Journal of Computer Vision, Vol. 88, No. 2, pp. 169-188, June 2010.

R. M. Karp. Implicit Hitting Set Problems and Multi-Genome Alignment. Keynote speech at the 21st Annual Symposium on Combinatorial Pattern Matching (CPM2010), New York, New York, June 2010.

B. Kirkpatrick, E. Halperin, and R. M. Karp. Haplotype Inference in Complex Pedigrees. Journal of Computational Biology, Vol. 17, No. 3, pp. 281-308, March 2010.

O. Kramer and H. Danielsiek. DBSCAN-Based Multi-Objective Niching to Approximate Equivalent Pareto-Subsets. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2010), Portland, Oregon, pp. 503-510, July 2010.

O. Kramer, B. Satzger, and J. Lässig. Power Prediction in Smart Grids with Evolutionary Local Kernel Regression. Proceedings of the Fifth International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010), San Sebastian, Spain, pp. 262-269, June 2010.

O. Kramer, B. Satzger, and J. Lässig. Managing Energy in a Virtual Power Plant Using Learning Classifier Systems. Proceedings of the 2010 International Conference on Genetic and Evolutionary Methods (GEM 2010), Las Vegas, Nevada, pp. 111-117, July 2010.

C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating Edge Network Neutrality, Security, and Performance. ICSI Technical Report TR-10-006, May 2010.

B. Kulis and P. Bartlett. Implicit Online Learning. Proceedings of the 27th International Conference on Machine Learning (ICML 2010), Haifa, Israel, June 2010.

J. Lässig and D. Sudholt. The Benefit of Migration in Parallel Evolutionary Algorithms. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2010), Portland, Oregon, pp. 1105-1112, July 2010.

J. Lässig and U. Trommler. New Approaches to Enterprise Cooperation Generation and Management. Proceedings of the 12th International Conference on Enterprise Information Systems (ICEIS 2010), Funchal, Portugal, June 2010.

K. Lee and D. Ellis. Audio-Based Semantic Concept Classification for Consumer Video. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, Issue 6, pp. 1406-1416, August 2010.

B. Nechaev, M. Allman, V. Paxson, and A. Gurtov. A Preliminary Analysis of TCP Performance in an Enterprise Network. Proceedings of the 2010 Internet Network Management Workshop/Workshop on Research and Enterprise Networking (INM/WREN '10), San Jose, California, April 2010.

F. Neumann, D. Sudholt, and C. Witt. A Few Ants Are Enough: ACO with Iteration-Best Update. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2010), Portland, Oregon, pp. 63-70, July 2010.

H. Pan, B. Hindman, and K. Asanovic. Composing Parallel Software Efficiently with Lither. Proceedings of the Programming Language Design and Implementation (PLDI 2010), Toronto, Canada, pp. 376-387, June 2010.

B. Pasaniuc, N. Zaitlen, and E. Halperin. Accurate Estimation of Expression Levels of Homologous Genes in RNA-Seq Experiments. Proceedings of the Fourteenth International Conference on Research in Computational Biology (RECOMB 2010), Lisbon, Portugal, pp. 397-409, April 2010.

S. Pötzsch and R. Böhme. The Role of Soft Information in Trust Building: Evidence from Online Social Lending. Proceedings of the Third International Conference on Trust and Trustworthy Computing (TRUST 2010), Berlin, Germany, pp. 381-395, June 2010.

J. Ruppenhofer, C. Sporleder, R. Morante, C. Baker, and M. Palmer. SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. Proceedings of the Fifth International Workshop on Semantic Evaluation (SemEval-2010) at the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, pp. 45-50, July 2010.

M. Salzmann, C. H. Ek, R. Urtasun, and T. Darrell. FOLS: Factorized Orthogonal Latent Spaces. Presented at the Learning Workshop, Snowbird, Utah, April 2010.

M. Salzmann, C. H. Ek, R. Urtasun, and T. Darrell. Factorized Orthogonal Latent Spaces. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010), Sardinia, Italy, pp. 701-708, May 2010.

M. Salzmann and R. Urtasun. A Constrained Combination of Discriminative and Generative Methods. Presented at the Learning Workshop, Snowbird, Utah, April 2010.

M. Salzmann and R. Urtasun. Combining Discriminative and Generative Methods for 3D Deformable Surface and Articulated Pose Reconstruction. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, California, pp. 647-654, June 2010.

S. Sankararaman, G. Obozinski, M. I. Jordan, and E. Halperin. Genomic Privacy and Limits of Individual Detection in a Pool. Presented at the 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2010), Boston, Massachusetts, July 2010.

B. Satzger, F. Bagci, and T. Ungerer. A Novel Constraint Satisfaction Problem Solver for Self-Configuring Distributed Systems with Highly Dynamic Behavior. Proceedings of the IEEE World Congress on Computational Intelligence (WCCI 2010), Barcelona, Spain, June 2010.

B. Satzger and O. Kramer. Learning Heuristic Functions for State-Space Planing. Proceedings of the Fifth IASTED International Conference on Computational Intelligence (CI 2010), Maui, Hawaii, August 2010.

B. Satzger, O. Kramer and J. Laessig. Adaptive Heuristic Estimates for Automated Planning Using Regression. proceedings of the 2010 International Conference on Artificial Intelligence (ICAI 2010), Las Vegas, Nevada, pp. 576-581, July 2010.

D. Skocaj, M. Kristan, A. Vrecko, A. Leonardis, M. Fritz, M. Stark, B. Schiele, S. Hongeng, and J. L. Wyatt. Multi-Modal Learning. In Cognitive Systems, H. I. Christensen, G.-J. Kruijff, A. Sloman, and J. Wyatt, eds., Springer, pp. 265-309, 2010.

R. Sommer and V. Paxson. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. Proceedings of the IEEE Symposium on Security and Privacy 2010, Oakland, California, May 2010.

R. Srivas, G. Hannum, A. Guénolé, H. van Attikum, N. J. Krogan, R. M. Karp, and T. Ideker. Genome-Wide Association Data Reveal a Global Map of Genetic Interactions among Protein Complexes. Presented at the 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2010), Boston, Massachusetts, July 2010.

L. Stoll and G. Doddington. Hunting for Wolves in Speaker Recognition. Proceedings of the Speaker and Language Recognition Workshop (Odyssey 2010), Brno, Czech Republic, pp. 159-164, June 2010.

Z. Tan, A. Waterman, R. Avizienis, Y. Lee, H. Cook, D. Patterson, and K. Asanovic. RAMP Gold: An FPGA-Based Architecture Simulator for Multiprocessors. Proceedings of the 47th Design Automation Conference (DAC 2010), Anaheim, California, June 2010.

Z. Tan, A. Waterman, H. Cook, S. Bird, K. Asanovic, and D. Patterson. A Case for FAME: FPGA Architecture Model Execution. Proceedings of the 37th International Symposium on Computer Architecture (ISCA 2010), Saint-Malo, France, pp. 290-301, June 2010.

G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tür, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, and F. Yang. The CALO Meeting Assistant System. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, Issue 6, pp. 1601-1611, August 2010.

M. Walfish, M. Vutukuru, H. Balakrishnan, D. Karger, and S. Shenker. DDoS Defense by Offense. ACM Transactions on Computer Systems, Vol. 28, Issue 1, No. 3, March 2010.

N. Weaver. Unencumbered by Success: The Usenix Security Grand Challenge Competition. ICSI Technical Report TR-10-004, April 2010.

N. Weaver, C. Kreibich, and V. Paxson. Measuring Access Connectivity Characteristics with Netalyzr. Presented at the meeting of the North American Network Operators' Group (NANOG49), San Francisco, California, June 2010.

T. Wittkop, D. Emig, S. J. Lange, S. Rahmann, M. Albrecht, J. H. Morris, S. Boecker, J. Stoye, and J. Baumbach. Partitioning Biological Data with Transitivity Clustering. Nature Methods, Vol. 7, Issue 6, pp. 419-420, June 2010.

N. Yosef, J. Gramm, Q.-F. Wang, W. S. Noble, R. M. Karp, and R. Sharan. Prediction of Phenotype Information from Genotype Data. Communications in Information and Systems, Vol. 10, No. 2, pp. 99-114, May 2010.

K. Zhou, L. Mo, P. Kay, V. P. Y. Kwok, T. N. M. Ip, and L. H. Tan. Newly Trained Lexical Categories Produce Lateralized Categorical Perception of Color. Proceedings of the National Academy of Sciences of the United States of America, Vol. 107, No. 22, pp. 9974-9978, June 2010.

INTERNATIONAL
COMPUTER SCIENCE
INSTITUTE

THE INTERNATIONAL COMPUTER SCIENCE INSTITUTE (ICSI) is the only independent, non-profit US lab solely conducting open, non-proprietary, pre-competitive research in computer science. Affiliated with the University of California campus in Berkeley, ICSI's mission is to further research in computer science through international collaboration, and further international collaboration through research in computer science. ICSI provides a haven for computer scientists to conduct long-term research without commercial limitations and with few faculty pressures. ICSI has significant efforts in these major research areas: Internet research, including Internet architecture, related theoretical questions, and network services and applications; theoretical computer science, including applications to bioinformatics; artificial intelligence, particularly for applications to natural language understanding; natural speech processing; computer architecture; and computer vision.

The ICSI Gazette is published twice a year by the International Computer Science Institute (ICSI).

INTERNATIONAL
COMPUTER
SCIENCE
INSTITUTE

1947 Center Street Suite 600
Berkeley, CA 94704 USA
v 510 666 2900
www.icsi.berkeley.edu