



INSTITUTE FOR SOFTWARE RESEARCH / UNIVERSITY OF CALIFORNIA, IRVINE

**HOT RESEARCH**

### Reconciling Personalization with Privacy

Personalization has become a household term in the online consumer market. E-tailers, search engines, and social networks all want to tailor their services to their users. Many surveys show that Internet users appreciate a personalized experience. At the same time, businesses that personalize their services see increased profits. Around 20-30% of Amazon purchases and 60% of Netflix views are a result of personalized recommendations.

The word “personalization” however points to its Achilles’ heel: to benefit from a personalized experience, users need to disclose personal information. This poses a problem since consumers today are very concerned about their privacy and hesitant to disclose information about themselves.

The Q1 2012 TRUSTe Privacy Index shows that 90% of U.S. adults worry about their privacy online.

ISR professor **Alfred Kobsa** works on reconciling

the twofold benefits of personalization with Internet users’ privacy concerns. He thereby collaborates with industry, including Samsung US R&D Center, Ericsson Research, and Microsoft Research.

One line of his research is to study how users can be assisted in their disclosure decisions by means of informative justifications for revealing personal information. For instance, one can explain one or more of the following to users:

- *How is the information used to personalize the users’ experience?* This

### RESEARCH BRIEFS

Prof. **Bonnie Nardi** co-authored the book *Activity Theory in HCI: Fundamentals and Reflections* with Victor Kaptelinin. Published by Morgan & Claypool Publishers, the book is part of the Synthesis Lectures on Human-Centered Informatics series.

Prof. **Gloria Mark**, with Stephen Volda and Armand Cardello, received Honorable Mention for Best Paper at the ACM Conference on Human Factors in Computing Systems (CHI 2012) for their paper “A Pace Not Dictated by Electrons’: An Empirical Study of Work Without Email.” This work was mentioned in the popular media including *The New York Times*, NPR, the BBC, and *The Atlantic*.

Prof. **Paul Dourish** delivered a keynote address, “The Commodification of Location: Reconfiguring Social Relations in Locative Media,” at the 3rd Annual Conference of the Pan-American Mobilities Network and the Cosmobilities Network (Local and Mobile 2012) in March in Raleigh, NC.

Prof. **Alfred Kobsa** recently gave two presentations in Singapore. “Privacy-sensitive design of personalized systems” was given at National University of Singapore in March and “Computer Game Environments for Personalized Self-Managed Chronic Care” was given at the Institute for InfoComm Research (I2R) in February.

Prof. **André van der Hoek** is serving as Program Committee Co-Chair for the 36th International Conference on Software Engineering (ICSE 2014), to be held in Hyderabad, India.

*More Research Briefs on page 5.*

strategy is in line with the “Consumer Privacy Bill of Rights” that was recently proposed by the White House. This guideline calls on industry to define a set of practices that will give consumers more information and control of their privacy. Specifically, the framework suggests that “companies should provide clear descriptions of [...] why they need the data, how they will use it.”

- *What are the benefits of disclosure?* People engage in a “privacy calculus,”

trading off the perceived benefits of disclosure with its potential risks. Justifications describing the benefits may tip the scale in favor disclosure or non-disclosure.

- *What did others do?* Indicating how many other users have disclosed the requested information allows people to

eschew their individual privacy calculus, by conforming to the behavior of others.

Do these justifications indeed decrease perceived

privacy threats and increase disclosure? Recent work by Ph.D. student **Bart Knijnenburg** (A. Kobsa, advisor) shows

### IN THIS ISSUE

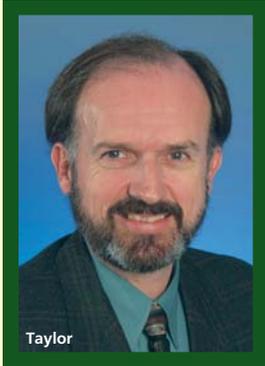
- 2-Message from the Director
- 3-Collaborations with Korea
- 4-Distinguished Alumnus Scacchi
- 4-Technical Reports
- 4-5-Student News
- 5-ISR Research Forum, Research Briefs
- 6-ISR Sponsorship, Student News
- 7-ISR at GSAW, Student News
- 7-Alumnus Owen O’Malley
- 8-ISR Events



Knijnenburg

## MESSAGE FROM THE DIRECTOR

In the parlance of the University of California system, the Institute for Software Research is an “Organized Research Unit,” or ORU. All ORUs are reviewed at five-year intervals by the University to determine if they are still performing top-tier research and providing important services to the University and to the community at large. ISR’s most recent review began last summer, with our preparation of a large summary-of-progress report, followed by scrutiny by anonymous reviewers external to the university, followed by still-further review by a committee within the university. After all that work (whew!) I am happy to report that we were renewed for another five years!



Taylor

On one hand I am not surprised, given that I am well-aware of the incredible breadth and depth of ISR-supported research and the degree to which we impact industry. On the other hand, as you are doubtless aware, the University of California system is under severe budget pressure due to California’s never-ending budget failures. Hence I am very pleased with the outcome of the review; it is a strong endorsement of our ongoing contributions. And “Thank you!” to all of those involved in the many stages of this review process.

On another note, I had the privilege of visiting Korea in March, presenting a two-day tutorial on software architecture at Samsung Electronics, and a one-day seminar on architecture at a software engineering professional society meeting in Seoul. While this was my first in-depth interaction with Korean colleagues, it was far from the first for ISR. As an accompanying article on Prof. **Walt Scacchi’s** visit to Korea (page 3) shows, there is significant focus in Korea on software engineering and software engineering education. Other ISR faculty have been engaged with Korea as well. The intersection of interests between ISR faculty and Korean industry is very strong, and I expect substantial

deepening of these ties in the future. Korean industry is committed to leveraging the best software engineering research results, and we look forward to working with them as partners in the near future. It is very gratifying to see such major industry players recognize the critical role of software engineering research in advancing their commercial interests.

On still another note, you are doubtless aware that we restarted our annual Research Forum in May. We had temporarily suspended this event due to budget pressures, but were able to reignite it this year. And I am so glad we did! The event was the most popular in our history, with a variety of faculty speakers, an open house for in-depth lab visits, and a keynote speaker, Professor Betty Cheng from Michigan State University. I was especially pleased to see many new faces from industry --- who had obviously done their homework and came looking for research that can help solve some of their problems. I heard reports of many excellent conversations,

and look forward to seeing those deepen into “meaningful relationships.” For more on the ISR Research Forum, see page 5.

Lastly, you may recall that in the previous issue of the *ISR Connector* I wrote about “Substance, Immediacy, and Focus,” commenting on social media, the tendency of our field to dutifully salute the latest buzzword, and the genuine need for consistent attention to “deep work.” I was somewhat surprised to hear back -- quickly! -- from many of you; the message must have resonated. Most responses were very much an endorsement of the message; one or two others not so much :-). Whatever your position I hope you will continue to join us in pushing forward the boundaries of our field.

ISR Director **Richard N. Taylor** can be reached at [taylor@uci.edu](mailto:taylor@uci.edu).

that the solution is not this simple. In an online user study with a mock-up of a mobile recommender system, Knijnenburg and Kobsa showed that users appreciate the disclosure help provided by the justifications. At the same time, though, these justifications decrease users’ trust, their satisfaction, and their

level of disclosure. This research has been described in ISR technical report UCI-ISR-12-1 which is available on the ISR website.

The likely explanation for this result is that the justifications tested by Knijnenburg and Kobsa did not only indicate high

disclosure benefits. Rather, promised benefits were sometimes less than stellar. In these cases, users not only withheld the information, but also became dissatisfied with the recommender for not providing better benefits. In fact, in-depth interviews revealed that users were much more attuned to the warning signal of low



## Seven Tips for Personalization with Fewer Privacy Problems

- Consumers are alarmingly hesitant to disclose their personal information online.
- Several national and numerous foreign laws regulate the collection of personal information. Many foreign laws also apply to U.S. sites if they collect personal data abroad.
- Consumers engage in a “privacy calculus”: they trade off the perceived benefits of disclosure with its perceived risks.
- Before requesting any personal information from users, one should ascertain that their benefits from its disclosure are justifiably large.
- Personalization can be a high-value benefit.
- Collecting personal information is a long-term trust-building exercise; simple persuasion techniques can cause a backlash.
- Giving users immediate feedback about the usefulness of their disclosure, and paying close attention to their individual privacy preferences and decision strategies, helps them in their privacy decisions and increases their satisfaction.

benefits than to the positive signal of high benefits. Hence, before requesting any personal information, system designers need to ascertain that users’ benefits from providing this information are high, since otherwise users are likely to rebel.

However, even when they justified the disclosure with very high benefits or very high social compliance, Knijnenburg and Kobsa found that users would still be less willing to disclose their personal information than when no justification was given. It might very well be that any justification carries an implicit warning, regardless of how positive it is. A justification inadvertently signals that the act of disclosure is not trivial and may involve risks. Users cannot be easily persuaded; collecting their personal information is a long-term trust-building exercise.

To solve these problems, Kobsa and his team are now looking into more sophisticated ways of justifying disclosure. One direction is the interactive

visualization of personalization benefits. Such visualization can explain the personalization process and give users immediate feedback about the usefulness of their disclosure, e.g., “since you told that you are a student we were able to recommend additional flights, which come with a student discount.” This research is carried out in collaboration with UC Santa Barbara.

Another direction is to tailor justifications to each user. For instance, preliminary results show that justifying disclosure by portraying the expected benefits actually works rather well for most users, except for males with low disclosure tendency. By paying close attention to people’s individual privacy preferences and decision strategies, system designers can give them a personalized experience without raising privacy concerns.

For more on Prof. Kobsa’s research, see:

<http://www.isr.uci.edu/~kobsa/>

Kobsa can be reached at [kobsa@uci.edu](mailto:kobsa@uci.edu).



Scacchi with Daegu City Mayor Kim Bumil

## Fostering Collaborations In Daegu, Korea

During the week of March 26-30 ISR Prof. **Walt Scacchi** visited with Daegu, Korea government officials, industry leaders, and academic colleagues to discuss establishment of a long-term research collaboration with ISR. The visits included a 30 minute meeting with the Mayor of Daegu City, Kim Bumil, who conveyed his personal interest in recognizing the win-win value of a long-term research relationship between Daegu and UCI. Daegu City has been designated by the Korean national government as the IT Convergence Center for South Korea, which will focus on research, development, and commercialization of software-based systems for advanced health care and medical research, green energy, advanced manufacturing, embedded software applications for consumer electronics and automotive systems, and computer games and 3D virtual worlds.

Dr. Scacchi also presented an invited lecture on the future of research in computer games and virtual worlds that was simultaneously translated to a standing room only audience of more than 150 participants at the Digital Industry Promotion (DIP) Agency headquarters in Daegu. Previously DIP has supported research of Scacchi and others at ISR, including Professor **Robert Nideffer**, during 2007-2010. Dr. Scacchi also met at length with the President of DIP, Jay Chae, to discuss potential new collaborative research projects between DIP and ISR. Among the most exciting of these possibilities is DIP’s interest in seeking to create a Software Convergence Center in a new multi-building complex, supported by the Korean national government. Such a center would include a

## ISR STUDENT NEWSBRIEFS

**Lee Martie** (A. van der Hoek, advisor) presented his paper "Trendy Bugs: Topic Trends in the Android Bug Reports" at the 9th Working Conference on Mining Software Repositories (MSR 2012), collocated with ICSE 2012 in Zurich Switzerland



in June. The paper is co-authored by Ph.D. students **Vijay Krishna Palepu** and **Hitesh Sajjani**, and Prof. **Cristina Videira Lopes**.

**Xinru Page** (A. Kobsa, advisor) presented her paper "Don't Disturb My Circles! Boundary Preservation Is at the Center of Location-Sharing Concerns" at the 6th Int'l AAAI Conference on Weblogs and Social Media



(ICWSM-12), in Dublin, Ireland, in June. The paper is co-authored by her advisor Prof. **Alfred Kobsa** and Ph.D. student **Bart Knijnenburg**.

**Ruy Cervantes** (B. Nardi, advisor), with his co-author Prof. **Bonnie Nardi**, presented the paper "Building a Mexican Startup Culture Over the Weekend" at the International Conference on



Intercultural Collaboration (ICIC) in Bangalore, India, in March.

For more information on students:

<http://www.isr.uci.edu/people.html>

building dedicated to software research, and another building hosting dozens of new or existing software companies focused on developing new applications informed by research knowledge produced or brought into their software research institute. Mr. Chae expressed great interest in seeing ISR as a model and lead collaboration partner in establishing this new software research institute, as well as in creating new collaborative software research projects between Daegu and UCI. Accordingly, ISR is poised to work with our friends in Daegu and elsewhere in South Korea to create international partnerships and collaborative research projects that can mutually benefit the participating parties.

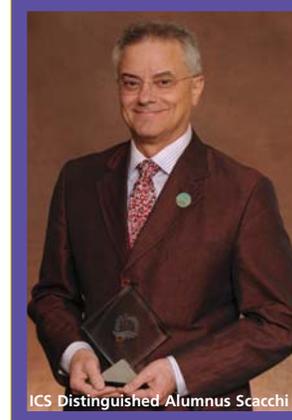
For more information on Scacchi and his research, see:

<http://www.isr.uci.edu/~wscacchi/>

Contact Scacchi at: [wscacchi@ics.uci.edu](mailto:wscacchi@ics.uci.edu).

## Walt Scacchi Honored with ICS Distinguished Alumnus Award

ISR research professor **Walt Scacchi** (Ph.D. '81), received the Donald Bren



School of ICS's 2012 Distinguished Alumnus Award at the 42nd annual Lauds & Laurels ceremony in May.

Dating back to his time as a graduate student, Scacchi has conducted nearly 40

years of research that has contributed significantly to the study of open source

## ISR Technical Reports Available Online

ISR technical reports present information resulting from student and faculty research carried out under the auspices of the Institute. They showcase early results not available in print elsewhere. ISR technical reports are available in PDF on the ISR website. Recent reports include:

"HealthWatch: A Diabetes Self-Management System for Personal Behavioral Rule Discovery"

Tao Wang  
UCI-ISR-12-7, June 2012

"Usability Inspection Method-based Analysis of a Socio-Technical Visualization Tool"

Erik H. Trainer, Stephen Quirk, Cleidson de Souza, David F. Redmiles  
UCI-ISR-12-6, June 2012

"Architecture-Driven Modeling of Adaptive Collaboration Structures in Large-Scale Social Web Applications"

Christoph Dorn, Richard N. Taylor  
UCI-ISR-12-5, May 2012

"Experimental Materials Used in the Study on Inspectability and Control in Social Recommender Systems"

Bart Piet Knijnenburg, Nikhil Rao, Alfred Kobsa  
UCI-ISR-12-4, April 2012

"Motile: Mobile Code for Decentralized On-Demand Tailored Services"

Michael Gorlick, Kyle Strasser, Richard N. Taylor  
UCI-ISR-12-3, April 2012

All ISR technical reports are available at:

<http://www.isr.uci.edu/tech-reports.html>

software development, large-scale software engineering and systems integration, and computer game culture and technology. His research combines technical innovation with a keen understanding of the important role that social factors play in the adoption of new technologies. Scacchi's research record includes more than 150 refereed publications. His projects have received support from a wide range of sponsors, including Digital Industry Promotion Agency of Daegu, Korea, Discovery Science Center in Santa Ana, CA, Intel Research Laboratories, Naval Postgraduate School in Monterey, CA, Northrop Grumman, and the San Francisco Symphony. In addition, he regularly serves as principal investigator on National Science Foundation grants, including a \$3 million four-year award received in 2008 to investigate a socio-technical approach to developing and using virtual activity systems.

Scacchi earned undergraduate degrees in computer science and mathematics at Cal State Fullerton and then joined UC Irvine's graduate program in Information and Computer Science. He completed his Ph.D. degree in 1981 and accepted a faculty position at the University of Southern California where he spent 17 years before returning to UCI in 1999.

Scacchi reflected: "It is an extraordinary honor to be recognized with this award by my friends and colleagues in the School of Information and Computer Sciences at UCI."

ISR congratulates Dr. Scacchi on his award!



## Industry and Academia Connect at the ISR Research Forum

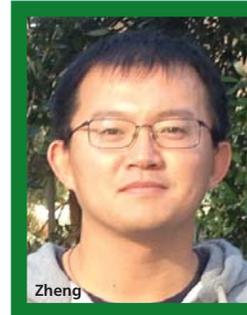
This spring saw the re-launch of the ISR Research Forum, a one-day event designed to encourage industry-academia interactions, collaborative research, and technology transition. Revamped with a new format featuring an 'open house' and quick 20 minute faculty research talks, the 2012 Forum was an enthusiastic success!

Industry attendees were drawn from a wide range of companies – from commercial software development, to consulting firms, to defense contractors and FFRDCs – including Boeing, Emida, Google, Health Focus, InfoLogic, JPL, MSC Software, Northrop Grumman, Samsung, The Aerospace Corporation, and more.

The program featured eight faculty talks providing an introduction to research of strategic import to industry. Topics included: the COAST architectural style, Computer Games and Virtual Worlds,

## ISR STUDENT NEWSBRIEFS

**Yongjie Zheng** (R. Taylor, advisor) gave two presentations at the 34th Int'l Conf. on Software Engineering (ICSE 2012), in June in Zurich. The first was a paper in the Research track titled: "Enhancing Architecture-Implementation Conformance with Change Management



and Support for Behavioral Mapping." The second, "xMapper: An Architecture-Implementation Mapping Tool", was an Informal Research Demonstration. Both are co-authored by ISR Director **Richard N. Taylor**. Zheng was also the recipient of a SIGSOFT travel scholarship to attend the ACM Turing Centenary Celebration in San Francisco in June.

**Lily Irani** (P. Dourish, advisor) gave a presentation titled "Designing Citizenship: Innovation, Empathy, and Social Reform in Elite India" at Stanford's South Asia by the Bay Graduate



Conference in May.

**Nicolas Lopez-Giraldo** (A. van der Hoek, advisor) is spending his summer as an intern at Google in Irvine, CA. He will be working on the Google Analytics API team under the supervision of Technical Engineering Lead Jeetendra Soneja.



## RESEARCH BRIEFS

Prof. **Hadar Ziv** gave a talk titled "Human Centered Software Engineering: If it doesn't work for people, it doesn't work!" at The Aerospace Corporation in El Segundo, CA, in June, as part of their Computer Science Division Tech Forum meeting series.

Prof. **Walt Scacchi** gave a keynote address titled "Game Development Competitions: Software Engineering as a Team Sport" at the Workshop on Games and Software Engineering at the International Conference on Software Engineering (ICSE 2012), Zurich, Switzerland in June.

ISR Visiting Researcher **Christoph Dorn** presented two conference papers, both co-authored by ISR Director **Richard N. Taylor**. "Analyzing Runtime Adaptability of Collaboration Patterns" was presented at the Int'l Conf. on Collaboration Technologies and Systems (CTS 2012) in May in Denver, CO, and "Co-Adapting Collaborations and Software Architectures" was presented at the Int'l Conf. on Software Engineering (ICSE 2012) – NIER Track in June in Zurich, Switzerland.

## ISR STUDENT NEWSBRIEFS

**Hitesh Sajnani** (C. Lopes, advisor) gave two presentations at the Int'l Conf. on Program Comprehension (ICPC) in Passau, Germany in June. The first, a paper in the student



research symposium titled "Automatic Software Architecture Recovery: A Machine Learning Approach," is co-authored by Prof. **Cristina Videira Lopes**.

The second, a poster titled "Parallel Code Clone Detection Using MapReduce," is co-authored by Ph.D. student **Joel Ossher** and Prof. **Lopes**. Sajnani was also the recipient of a SIGSOFT travel scholarship to attend the ACM Turing Centenary Celebration in June.

**Francisco Servant** (J. Jones, advisor) presented his paper "WhoseFault:



Automatic Developer-to-Fault Assignment through Fault Localization" at the 34th Int'l Conf. on Software Engineering (ICSE 2012), in Zurich,

Switzerland in June. The paper is co-authored by Prof. **James A. Jones**.

**Alegria Baquero** (R. Taylor, advisor) is interning this summer at



CommerceNet in Palo Alto, CA where she will focus on privacy and medical records/systems. She will be working with Exec. Dir. Allan Schiffman.

## WANT TO GET INVOLVED?

Sponsoring ISR has many benefits. It enables your company to form closer ties with our faculty and students, puts you on the fast track to our leading edge research, and gives you first crack at our experimental software tools. Choose from five levels of sponsorship:

| Support Level          | Annual Contribution | Contribution goes to:           |
|------------------------|---------------------|---------------------------------|
| Friend                 | \$10,000            | ISR's general research fund.    |
| Affiliate—Research     | \$30,000            | A designated ISR research area. |
| Affiliate—Visiting     | \$40,000            | Faculty host's research area.   |
| Affiliate—Grad Student | \$60,000            | Graduate student fellowship.    |
| Partner                | \$100,000 or more   | Large-scale research project.   |

For more information about ISR Sponsorship, please contact:

**Debra A. Brodbeck**  
brodbeck@uci.edu  
(949) 824-2260

Runtime Bloat Analysis, Federations, Architectural Degradation, Software Sketching and Early Design, Socio-technical Dependencies and Software Architecture, and Managing Digital Information in the Workplace.

The clear-cut hit of the day was the open house, new this year. Attendees were able to visit faculty research labs in multiple buildings, view research demonstrations in an interactive environment, and engage students and researchers directly as they presented their work via posters.

Prof. **André van der Hoek** remarked: "The Open House was a highlight, with up-close and personal interactions between ISR students and faculty and attendees about our research and its potential in industry. It was particularly gratifying to see attendees try out the Calico whiteboard sketching software and the Code Orb themselves, showing we just may be reaching ISR's objective of producing research of importance to industry and that can be transitioned to actual use."

The event closed with an engaging keynote talk on "Model-based Development of High-Assurance Dynamically Adaptive Systems" by guest speaker Prof. Betty H.C. Cheng of Michigan State University.

For more information, including presenters' slides and a video of the keynote, visit:

<http://www.isr.uci.edu/events/Research-Forum-2012/>

## Become Part of the ISR Family

Rubbing elbows with ISR faculty, staff and students gives you a valuable window into the technology landscape of the future. But a relationship with ISR can be much more: Think of us as an extension of your company — a think tank, an R&D department, a research library, a consulting



firm, a training department, and an employment agency, all rolled into one. More importantly, when you sponsor ISR you become part of a friendly group of folks who speak the

same language and are eager to work with you to solve your current technical problems in the most cost-effective way possible.

Be part of the ISR Family — a Friend, Affiliate, or Partner.

For more information, visit:

<http://www.isr.uci.edu/sponsorship.html>  
or contact:

**Debra A. Brodbeck**  
Technical Relations Director  
brodbeck@uci.edu, (949) 824-2260

## ISR Participates in 2012 Ground System Architectures Workshop (GSAW)

On February 27, ISR Prof. **Walt Scacchi** delivered a half-day tutorial at GSAW 2012 on the topic of “Using Open Source Software in Ground Systems.” The tutorial addressed the state of the art in open source software development (OSSD) processes, work practices, and project

community dynamics, and examined results from recent empirical studies of OSSD. It was targeted towards software developers, system architects, project managers, and others.

The well-attended tutorial included participants from a mix of aerospace companies and government agencies.

Lively discussion ensued both during and after the tutorial, regarding such topics as: the roles open source software and OS development processes can play in support of the development of future ground systems; different open source development models, and the circumstances in which they succeed; and the kinds of benefits open source contributors achieve – including contributors from the defense arena. According to Scacchi, “Participating in GSAW provides many benefits to ISR, including the opportunity to interact with industry and government agency colleagues, to ground our research, and to explore new collaborative research opportunities.” Scacchi previously gave a tutorial on OSSD at GSAW in 2009, part of ISR’s lengthy tradition of GSAW participation.

The annual Ground System Architectures Workshop is sponsored by The Aerospace Corporation – one of ISR’s long-time supporters. GSAW has been held in cooperation with ISR since 2003. GSAW 2012 was held February 27 – March 1 in Los Angeles.



### ALUMNI PROFILE

## Dr. Owen O'Malley, 1996 Graduate

When **Owen O'Malley** was picking graduate schools, he chose UC Irvine’s Information & Computer Science because of its leading-edge software group and strong ties to industry. In software, it is critical to understand your customer and their use cases. Doing research is very exciting, but it is critical to combine the theoretical with the experience of industrial professionals. At UCI, with the guidance of his advisor Prof. **Debra J. Richardson**, he explored many different topics that proved invaluable to him in his career including program analysis, graph algorithms, distributed computing, and software architectures.

The Monday after turning in his dissertation, O'Malley started at Reasoning, which was a 10 year old start up that had just pivoted to produce a tool to find and fix y2k problems in legacy COBOL applications. “The experience I had at UCI in static analysis and developing data flow tools were critical in that job. But I never expected to build a data analysis tool for COBOL!” said O'Malley.

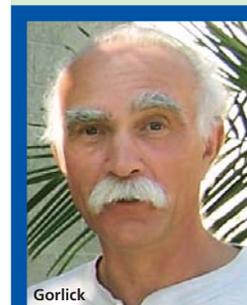
After saving the world from y2k disasters, O'Malley moved to Sun where he worked on a team that developed the revision control system that managed the data used for designing Sun’s CPUs. “The project was challenging because we had three sites working on the CPU designs and the data was too big to access from a single central server. The data had to be distributed and had to be kept in sync between the three sites even when they were generating 100,000 new file versions in a single day,” said O'Malley.

### ISR STUDENT NEWSBRIEFS

**Nicholas Mangano's** (A. van der Hoek, advisor) paper, “The design and evaluation of a tool to support software designers at the white-board,” has been accepted to the *Journal of Automated Software Engineering*. The paper is co-authored by his advisor Prof. **André van der Hoek**.



**Michael Gorlick's** (R. Taylor, advisor) paper “COAST: Architectures for



Decentralized On-Demand Tailored Services” has been accepted to the 10th Working IEEE/IFIP Conf. on Software Architecture (WICSA) to be held in

Helsinki, Finland in August. The paper is co-authored by Ph.D. student **Kyle Strasser** and their advisor ISR Director **Richard N. Taylor**.

**Nicholas DiGiuseppe** (J. Jones, advisor) presented his paper “Software



Behavior and Failure Clustering: An Empirical Study of Fault Causality” at the Fifth IEEE International Conference on Software Testing, Verification

and Validation (ICST) in Montréal, Canada in April. The paper is co-authored by Prof. **James A. Jones**.

For more information on students:

<http://www.isr.uci.edu/people.html>



After Sun, O'Malley moved to NASA. O'Malley's parents met each other while working as programmers on NASA's Apollo project, so not only is he a second generation programmer, he has always been interested in NASA. While there, he worked on many exciting projects. For example, in 2002, he had the opportunity to work at NASA Ames Research Center on a software model checker named Java Path Finder. Software model checking is a form of static analysis that explores the possible states of a program looking for failures. While O'Malley was analyzing software that suggested actions in space shuttle emergencies, he got to pilot a three engine out scenario on one of the simulators that the shuttle pilots use for training and managed to land the simulated shuttle on the first try. O'Malley also helped out on the Mars Exploration Rover project. When the Rover had just landed on Mars, the planning software was crashing occasionally. As an expert in C++, O'Malley was called in to figure out the problem and he discovered that they were hitting a bug in the C++ runtime, which had a simple fix.

O'Malley joined Yahoo Search's WebMap team after NASA. WebMap built and analyzed a graph of the known web with a node for each URL and an edge for each link. It had 100 billion nodes and a trillion edges and the compressed graph was 100 terabytes. Needless to say, it took a lot of computers to build it in a timely manner. Although it had been scaled up to run on 800 computers, its framework needed to be replaced to support an even larger scale. The team started designing and prototyping a new C++ framework based on the GFS and MapReduce papers from Google, but discovered a similar distributed file system and MapReduce implementation in the Lucene project at the Apache Software Foundation. It only ran on 20 machines, but it had the huge advantage of already being open source. O'Malley's team had planned to open source their new framework, but starting with an open source project made it easier. That code quickly became the Apache Hadoop project, which became the de facto standard for big data processing. Two other UCI alumni, Hairong Kuang and Koji Noguchi, were also early critical members of the Hadoop team. Yahoo has over 45,000 computers running Hadoop; the largest clusters are 4,500 computers, and are critical to Yahoo's business. Because the project is open source and is unique in the scale of

## ISR EVENT SCHEDULE

*Mark your calendars now!*

May 31, 2013

### ISR Research Forum: Where Research Meets the Real World

The 2012-2013 ISR Distinguished Speaker Series will be announced in Fall.

We look forward to seeing you!

For more information:

<http://www.isr.uci.edu/events.html>

data it can handle, it has been adopted by numerous companies including Facebook, LinkedIn, eBay, Apple, and Twitter.

In 2009 O'Malley and his colleagues used Hadoop to set the world record for big data sorting. They used 1,406 machines to sort a terabyte of data (10 billion records) in just 62 seconds and 3,658 machines to sort a petabyte of data (10 trillion records) in only 16.25 hours.

The power of Hadoop is that it lets you use many computers together, but the problem of using many computers is that they are always breaking. Computers typically last 3 years, but that means for every 1000 computers you'll lose one a day. If you use the computers heavily, you'll lose several each day. To keep Hadoop usable, it automatically handles the failures in software. A story O'Malley shared demonstrates how this works. The operator of the Hadoop clusters at LinkedIn was trying to convince his management that they didn't need to buy computers with redundant power supplies, even though that's what they typically did. In the meeting, the operator pulled up the list of machines in their production cluster and asked the managers to select a machine. The operator logged in to the one they chose and halted the machine with no warning or shutdown. Hadoop keeps three replicas of the information, so when a computer goes down the missing data is copied from one of the remaining replicas. Equivalently, the compute jobs on that computer are

automatically reassigned to other computers. As a result, operators don't need to fix problems immediately, and one operator can run 3,000 computers. So in the operator's 'experiment', the production cluster continued to function successfully.

Last year, Benchmark Capital approached Yahoo saying there was a great business opportunity to spin out the Hadoop group into a separate company. As a result, in July 2011, O'Malley and his co-founders started Hortonworks with 25 people. Hortonworks has quickly grown over the last year to 80 people. O'Malley said, "It's really exciting to be paid to work on open source full time. Further, Hortonworks' commitment to open source means that all our work gets released as open source, which is rewarding." Since their software is available for free, Hortonworks' revenue is generated via training and support.

Dr. O'Malley can be reached at [omalley@apache.org](mailto:omalley@apache.org).

# ISR

To receive the ISR Connector, send an email request to: [isr@uci.edu](mailto:isr@uci.edu)

ISR news, including the ISR Connector, is available at the ISR website: <http://www.isr.uci.edu/>

For more information, contact:

Debra A. Brodbeck  
Technical Relations Director  
[brodbeck@uci.edu](mailto:brodbeck@uci.edu), (949) 824-2260

UNIVERSITY of  
CALIFORNIA  
IRVINE

## SPECIAL THANKS

The UCI Institute for Software Research is generously supported by:

Northrop Grumman  
San Francisco Symphony  
The Aerospace Corporation

