UMIACS
UNIVERSITY OF MARYLAND INSTITUTE FOR ADVANCED COMPUTER STUDIES

InterConnections

January 2007

Contents

InterConnections is published by the University of Maryland Institute for Advanced Computer Studies (UMIACS).

Visit UMIACS on the web at: *www.umiacs.umd.edu*.

Director: V.S. Subrahmanian

Contact information
UMIACS
2119 A.V. Williams Bldg.
University of Maryland
College Park, MD
20742-3251
Phone: 301.405.6722

Adam Porter: Revolutionizing Globalized Software Development Processes

Professor Adam Porter

Software scale and complexity are growing by every measure: there is an ever increasing diversity of hardware, network components, software components, storage modules and data models. Moreover, these components are getting increasingly linked in bewildering and previously unforeseen ways. At the same time, business trends are squeezing development resources while demanding ever faster delivery times. To keep up, UMIACS and Computer Science professor Adam Porter finds that software development projects are struggling to become more "agile", using loosely coordinated, often globally distributed development teams to assemble highly configurable systems from generic services and components and then incrementally adding new features to the system over time.

Porter says he "wants to create the revolutionary new tools that will help companies succeed in this new environment". He believes that this kind of software development only works if you have rapid, highly effective and scalable software quality assurance. Otherwise, undetected problems accumulate, which eventually results in costly rework that erases the hoped-for efficiencies. "Unfortunately," he says, "existing quality assurance (QA) processes don't meet these standards."

Therefore, Porter has developed a new approach to QA called ***Distributed, Continuous Quality Assurance*** (***DCQA***). Inspired by volunteer computing projects such as SETI@Home and Folding@Home - which distribute pieces of enormous scientific computations to computers spread across the world - he has redesigned traditional QA analyses so that they can be efficiently run across extensive grids of computing resources, volunteered by worldwide developer and user communities, in a distributed and continuous manner. This approach greatly improves the quality and speed of QA processes, gives developers greatly expanded insight into system performance on diverse runtime environments and workloads, and allows efficient, coordinated and transparent execution of very large scale QA processes.

Take, for instance, the ACE+TAO+CIAO (ATC) project, which is a widely used suite of software that runs in thousands of commercial, academic and military systems, including DIRECTV satellite TV, the Keck observatory, and the USS Ronald Reagan aircraft carrier.

continued on page 2...

Director's message

UMIACS Director V.S. Subrahmanian

The second half of 2006 has seen tremendous advances by UMIACS faculty who have seen the growing impact of their work in the academic, corporate, and government sectors.

UMIACS and Computer Science Professor Dianne O'Leary leads the pack with recognition of her numerous accomplishments in numeric algorithms, scientific computation, and information retrieval – Dianne was named both an ACM Distinguished Scientist and an ACM Fellow in late 2006. UMIACS Senior Research Scientist Dave Doermann and his students came away with both a best student paper and a best paper award at a leading conference on handwriting analysis and mobile multimedia. Yiannis Aloimonos and his students won yet another best paper award at a leading computer vision conference. Not to be outdone, Bill Arbaugh and his student won a best paper award and Uzi Vishkin and his students both won best paper awards at conferences in their field, while Rama Chellappa and his students won a best student paper award at a vision conference. Professor Joseph JaJa (UMIACS/ECE) was part of one of the four winning teams of the first Annual Internet 2 IDEA (Internet2 Driving Exemplary Applications) Awards. Moreover, a team of UMIACS and CS students won the best paper award at the European Semantic Web conference. I am proud to report on this incredible period of academic recognition for our faculty and students. Congratulations to all!

UMIACS faculty have had significant impact outside of academia as well. The recently created Lab for Computational Cultural Dynamics (LCCD) received a 5-year $6 million award from the Air Force Office of Scientific Research to further study how to build models of the behaviors of different groups around the world. Concurrently with these advances, LCCD staff have extracted and delivered data on tribes in the Pakistan-Afghanistan borderlands to the US Army's 10th Mountain Division prior to their deployment in that part of the world. UMIACS and computer science Professor Steven Salzberg is a much sought after speaker in the news media for the pioneering work he is doing on identifying the genetic structure of influenza related organisms.

Last, but not least, I am pleased to welcome Professor Jonathan Wilkenfeld from the Department of Government and Politics to UMIACS. Jon is a pioneer in the study of terrorist groups worldwide, as well as "pre-terrorist" groups. He runs the "Minorities at Risk" project that has been gathering data on several hundred of these groups for many years – he will be working with UMIACS faculty in LCCD to develop models of the behaviors of these groups. Welcome on board, Jon!

Porter

...continued from cover

The systems comprise over 2 million lines of code that run across dozens of OS and compiler combinations. It is highly configurable, with over 500 configuration options. It is maintained by about 40 developers geographically distributed around the world who change the code frequently, checking in over 400 updates per week on the average. It can take anywhere from 3 to 18 hours to build/test just one system configuration, depending on the machine used. As a result, individual ATC developers can't test their changes on all possible platform and operating system combinations because they don't have adequate access to diverse operating systems, compilers, hardware platforms, CPU cycles, or disk space to run the hundreds of system regression tests. Moreover, given the several hundred orthogonal features/options that can be enabled/disabled, testing all possible configurations goes far beyond the resources of the core ATC development team. The end result is that ATC is forced to release code that's never been tested, and struggles with huge blind spots about how their systems actually behave in the field, how their design decisions affect such behavior, and how to interpret reports of failure arriving from the field.

To improve this situation, Porter and his colleagues developed several new DCQA processes which they applied to ATC. For example, one process helps ATC developers quickly identify specific errors and characterize in which configurations those errors occur. It is done by modeling the space of all system configurations, field testing specific configurations, and then feeding the results to machine learning algorithms.

Because there are many configurations--each taking hours to test--they developed search techniques and mathematical sampling strategies to improve early fault characterization. This information allows ATC developers to quickly narrow down the causes of the specific failures.

In another example, Porter and his colleagues developed a DCQA process to help developers determine whether a recent system update has unacceptably degraded system performance. This process executes statistically-designed experiments across a volunteered computing grid. The results are then analyzed to identify a small subset of the most important performance-related configuration options. Whenever software changes occur thereafter, developers can quickly estimate system performance across the entire configuration space by exhaustively exploring all combinations of the important options, while randomizing the rest. Again, ATC developers can use this process to cheaply, rapidly and reliably iden-

Porter's DCQA model is used to verify ATC software that is deployed on ships such as the USS Ronald Reagan (shown above). Courtesy: defenselink.mil

tify when software changes cause unexpected performance degradation.

In a third example, applied to a different project, Porter and colleagues at the National Institute of Statistical Sciences and Georgia Tech developed an approach that intelligently captures and analyzes execution data taken from many fielded program instances to find clues as to the root causes of system failures. Unlike other approaches which require each program instance to collect the same, expansive set of data (and thus suffer significant space, performance, and analysis overheads) their approach activates a different small subset of measurement probes in each instance. To process this data, they then developed a new classification technique called Association Trees that can build classification models even when the different data sets (program instances) have different predictors (measurement locations activated). Later, they extended this approach with a sequential version called Adaptive Sampling Association Trees, which incrementally learns the best predictors and then preferentially activates them in future instances. Their empirical results suggest that the approach is as effective as existing techniques, but requires only a small fraction of the data.

Backed by these successes, Porter and colleagues at UM and Vanderbilt recently won an equipment grant from the Office of Naval Research (ONR) Defense University Research Infrastructure Program (DURIP) to develop a large-scale distributed evaluation testbed. This testbed consists of a pair of dedicated clusters containing over 225 top-end x86 CPUs that can be individually configured to run different operating systems, support software and libraries. Professor Porter's successful DCQA effort has led to wide industry interest leading to collaborations with several high profile open source projects such as MySQL and JBoss Messaging, as well as leading edge corporations such as Cisco, Raytheon and Lockheed Martin.

News in Brief

Yiannis Aloimonos's research and commentary on Geoffrey Hinton's work were featured in the article "Neural Networks Show New Promise for Machine Vision" in IEEE Magazine for Computing in Science and Engineering.

William Arbaugh and T. Charles Clancy's paper entitled "Measuring Interference Temperature" submitted to 16th Annual Symposium on Wireless Personal Communications was selected as "Best Paper".

Ben Bederson, Director, HCIL, has won the prestigious IBM Faculty Award for his work on the Piccolo 2D graphics toolkit. The award, for 2006, is $30,000.

Rama Chellappa's research on characterizing human motion was featured in a January 5, 2007 Baltimore Sun article. (http://www.baltimoresun.com/news/health/bal-hs.gait05jan05,0,1884420.story)

Justin Domke and Yiannis Aloimonos won the best paper award at the Photogrammetric Computer Vision Conference that took place in Bonn, Germany, September 20-22, 2006.

Assistant Professor Amol Deshpande has received an NSF Career Award.

Lise Getoor and Indrajit Bhattacharyya won the best paper award at the SIAM Data Mining Conference.

UMIACS post-doc Jen Golbeck has been named one of the IEEE Intelligent Systems 10 to Watch. This is their first ever top 10 AI list and was featured in an article in the May/June 2006 special issue on the Future of AI, which commemorated the 50th Anniversary of the Dartmouth Workshop (generally considered the birthplace of modern AI).

UMIACS and CS students Aditya Kalyanpur, Bijan Parsia, Evren Sirin, and Bernardo Cuenca-Grau, won the Best Paper Award at the European Semantic Web Conference for their paper: "Repairing Unsatisfiable Concepts in OWL Ontologies."

Former UMIACS Director Joseph JaJa's project, "Transcontinental Persistent Archives Prototype" was among four winners of the first annual Internet2 Driving Exemplary Applications (IDEA) Awards. This program recognizes and encourages innovative advanced network applications that have had the most positive impact within the research and education community.

Dianne O'Leary has been named a Distinguished Scientist by the ACM in recognition for her outstanding contributions to computing. She has also been named a Fellow of the ACM.

V.S. Subrahmanian has been selected for inclusion on ISIHighlycited.com - a web site run by Thomson Scientific and ISI as part of their ISI Web of Science citation rankings. According to ISIHighlycited.com, "ISIHighlyCited.com will grow to include the top 250 preeminent individual researchers in each of 21 subject categories who have demonstrated great influence in their field as measured by citations to their work--the intellectual debt acknowledged by their colleagues."

OASYS, the Opinion Analysis System from UMIACS, developed by V.S. Subrahmanian, his students, Bonnie Dorr, and collaborators from the University of Naples, is one of the ten award-winning technologies for Computerworlds' 2006 Horizon Awards.

Rama Chellappa: Gait and Face Recognition Technology Could Move Video Interpretation to a New Level

Professor Rama Chellappa

Surveillance cameras are sprouting up in more and more places, forming an ever more powerful tool for solving crimes after they happen. But what about using them to prevent or stop criminal and terrorist acts? This requires that someone, or something, watch these rapidly multiplying video feeds 24-7.

And that's the problem. Paying people to adequately monitor dozens, or even hundreds, of surveillance cameras can be highly expensive. Plus, humans tend to get bored and lose focus staring at TV monitors hour after hour, day after day, watching for suspicious behavior that may occur only rarely, if ever. Computerized monitoring would seem to be the obvious answer, but creating software programs that can recognize suspicious activities or suspect individuals has proven highly difficult.

However, Rama Chellappa, who is a permanent member in the University of Maryland Institute for Advanced Computer Studies (UMIACS), a professor in the electrical and computer engineering department, and an affiliate professor in the department of computer science, has developed a real-time computer monitoring system that provides some answers to this problem. Chellappa's video interpretation system can reliably monitor surveillance images to detect certain suspicious movements or suspect individuals and alert human security personnel.

A pioneer in the development of pattern recognition and computer vision software, Chellappa and some of his associates demonstrated the system for military leaders at the 25th Army Science Conference. This event was the latest of numerous invited presentations on his system that Chellappa has made this past fall. In October, he delivered a plenary lecture on this emerging technology at the IEEE International Conference on Image Processing in Atlanta.

Chellappa's technology could have widespread applications in security surveillance, as well as non-security applications in elder care and video indexing. It could be particularly helpful as part of the security measures for sensitive locations, such as military bases, public transportation terminals and other areas that need high levels of security.

How It Works

Using video data from digital surveillance cameras and corresponding algorithms, Chellappa and his research assistants have developed a compact, digital signature for characterizing human gait and corresponding activities, such as humans carrying objects like backpacks, handbags, or briefcases.

When a person's limbs are unencumbered, gait movements are symmetrical. Represented graphically, these movements form a twisted helical pattern resembling a "figure 8" called a double helical signature. An individual's gait pattern is changed by any activity that changes the symmetry of the movements, for example, carrying a package. By defining these signatures, the system can recognize unique patterns in human gait and automatically detect asymmetric movement, such as an individual walking with a hidden object tied to an ankle or wrist.

Chellappa and his team have integrated human gait signature into a real-time video surveillance system and used it to study and locate pedestrians. The experimental results have demonstrated the effectiveness of the system under lighting changes, shadows, camera motion, various viewing angles, as well as significant obstacles in the cameras' viewing angles. The results also indicate that the approach is superior to many existing methods in terms of accuracy and reliability.

"These capabilities are extremely useful in creating a video interpretation system intended to address security concerns," said Chellappa.

His research team is also "teaching" their gait recognition system to identify individuals by their unique gait. This is a much more difficult task, since subjects may deliberately attempt to walk in an uncharacteristic manner

Illustration of automatically measuring the heights of humans as they move across the field of view of a camera. Data courtesy of Honeywell Corporation.

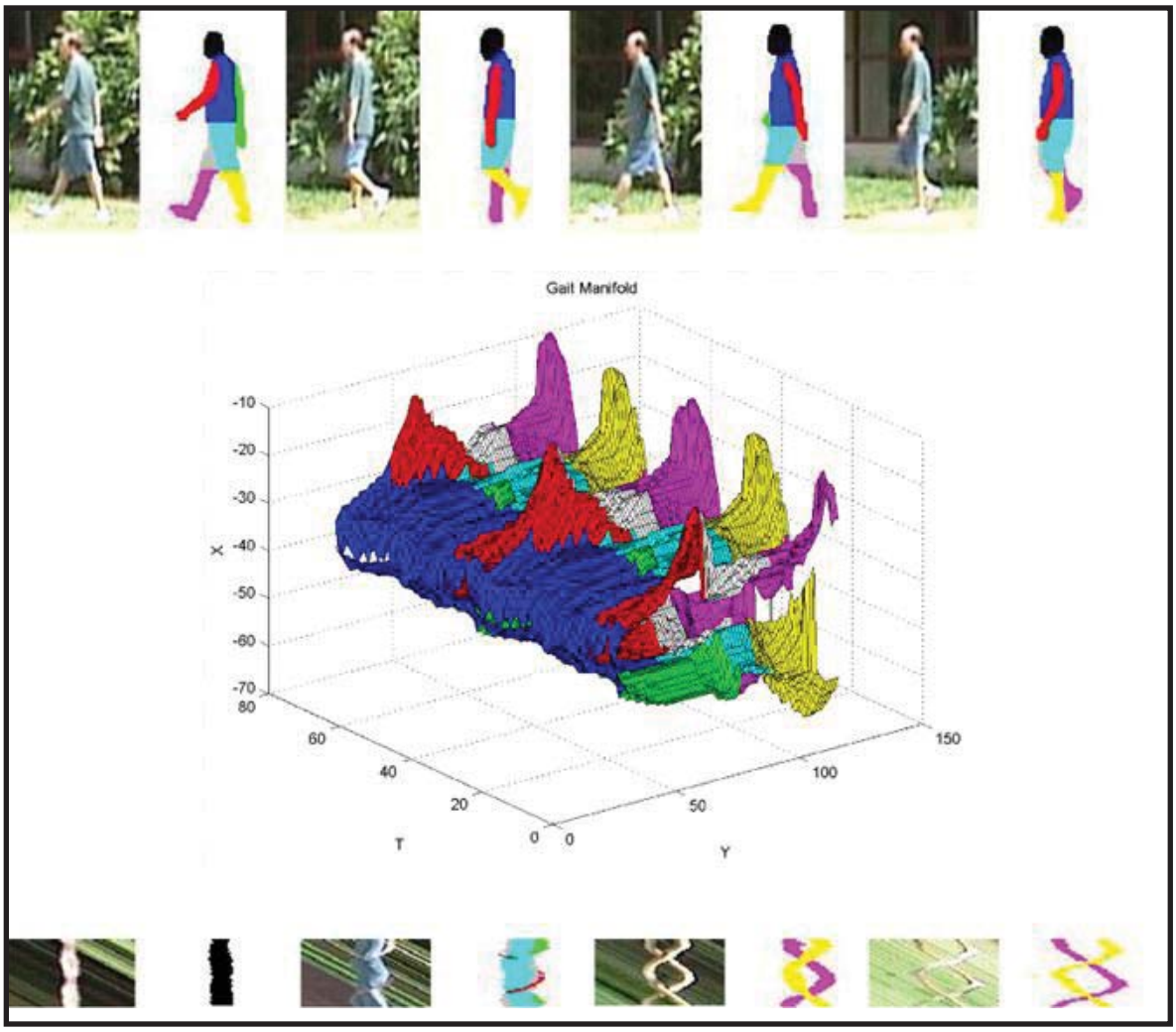


Illustration of double helical signature derived from a video sequence of a walking human. The first row shows a walking human whose body parts have been color-coded. The 3D gait manifold is shown in the center. The DHS generated by the walking human are shown in the bottom row. The DHS codes static and dynamic features of human gait. Reproduced from the Ph.D. dissertation of Dr. Yang Ran.

in order to try and cheat the system and avoid detection. If suspects are unaware of the surveillance system, their normal walking styles are more easily identified.

Chellappa and his assistants also study the geometric constraints that are useful in matching gait signatures across different viewing directions and individuals. The optimal camera angle for recognizing human gait is a sideways, 90-degree profile perspective, but Chellappa and his team have created automatic, corrective algorithms that can, within a certain range, compensate for different viewing angles.

Combining Gait, Face and Other Recognition Technologies

The Maryland team has also developed advanced face recognition software that can be combined with their gait recognition technology. This face recognition technology can be used to watch for known terrorists, spies or criminals and help to identify unknown individuals who might turn up repeatedly in sensitive locations or who have been present during multiple criminal or terrorist acts.

Chellappa's team recently developed two other recognition technologies that can add to the capabilities of automated video surveillance systems. Through work supported by the department of Homeland Security, Chellappa and one of his students have developed an algorithm to accurately estimate the heights of subjects in the field of view of a camera. This provides an important additional way to recognize and track subjects in crowded settings. The team has also developed a program that detects unattended packages using a structured representation known as attribute grammars. Both of these technologies were demonstrated at the recent Army Science Conference. The live demonstration of the height assessment program, in particular, attracted much interest and many volunteers, including many high-level army officers.

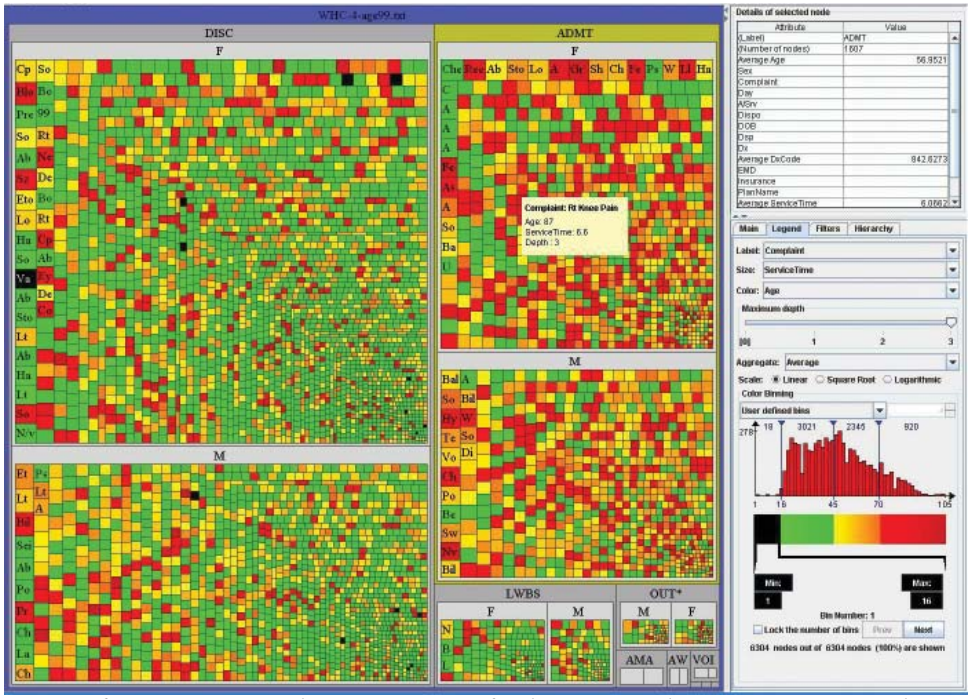
Chellappa and his team now are working to integrate their patent-pending technologies into a comprehensive surveillance system for use in security-sensitive locations.

From Invention to Application ...continued from back cover.

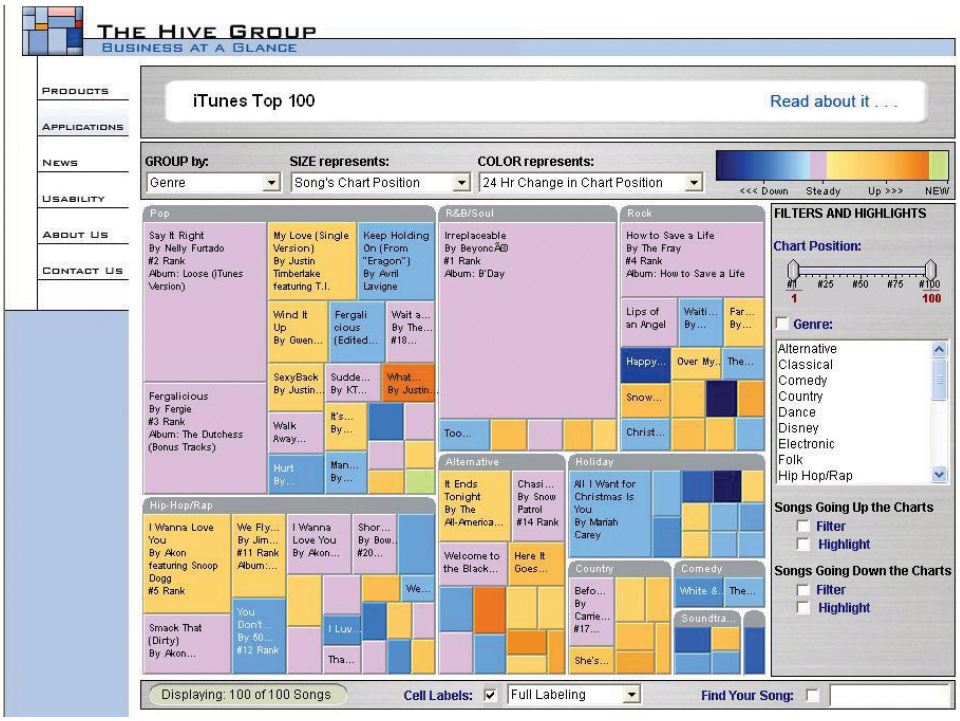
activity. Their updated software, with an Excel Add-In, remains available (http://research.microsoft.com/community/treemapper/).

Partnering with Wattenberg and a UMIACS colleague, Professor Ben Bederson, they developed still more advanced ideas that led to a new paper in the *ACM Transactions on Graphics* in 2002. Shneiderman recalls: "This is one of my all-time favorite research papers. It was fun to work with such clever colleagues and the paper had a terrific set of new results: an innovative algorithm, compelling visual displays, execution speed analysis, empirical studies with users, and new applications to photo libraries."

The University of Maryland Treemap software continued to be refined in partnership with Dr. Catherine Plaisant (UMIACS Associate Research Scientist). It has been available for free downloading, and 17 companies have acquired licenses through the Office of Technology Commercialization. The HiveGroup (www.hivegroup.com), a commercial provider of treemap solutions, has exclusively licensed the commercial interest of the UM version of Treemap software. The HiveGroup also developed a commercial application called Honeycomb which is widely used across manufacturing, logistics, defense, intelligence and financial services organizations.



Treemap of 6304 patients seen in the emergency room of Washington Hospital Center in January 2006. The reddish rectangles show that the discharged patients (DISC) were much younger (average age 47) than the admitted patients (ADMIT) (average age 57). Female patients outnumber male patients. Size of rectangles shows service time and the labels (where visible) are codes for symptoms, such as Cp for chest pain.

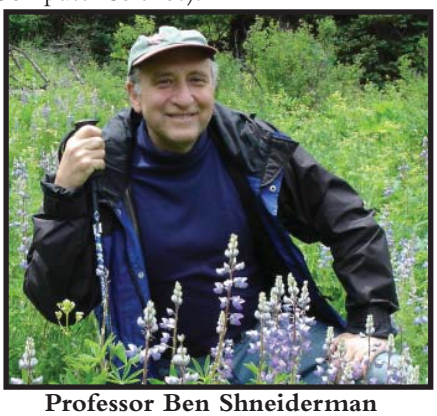


HiveGroup's demonstration treemap with data showing the Apple iTunes Top 100 songs. Size indicates ranking (#1 is biggest) and color indicates whether the song is rising or falling in the rankings. (www.hivegroup.com)

Competition is heating up among suppliers of treemap software. Recent applications have grown to serve thousands of users in diverse domains such as fraud detection in Korea, project management in Brazil, financial analysis in London, and world-wide supply chain management by the U.S. Marine Corps.

Shneiderman comments: "It's a satisfying story with exciting beginning, pause in the middle, and then a vigorous turnaround. I'm pleased that there are thousands of references to treemaps and dozens of implementations. It's great to see that some research ideas become widely applied."

Shneiderman's long history of contributions to the emergence of the discipline of Human-Computer Interaction will be celebrated later in 2007 with a special issue of the *International Journal of Human-Computer Interaction* that honors his 60th birthday. Appropriately, the special issue editors will be Shneiderman's research collaborator for 20 years, Dr. Catherine Plaisant and former PhD student Dr. Christopher North (Associate Professor, Virginia Tech Dept of Computer Science).



Professor Ben Shneiderman

Louiqa Raschid: Helping those Affected by Disasters

A large scale disaster brings together a diversity of people and organizations, and requires the discovery, modeling and integration of multiple, potentially complex, heterogeneous and rapidly evolving sources of data, to meet their information needs. According to UMIACS Director V.S. Subrahmanian, "the data systems and processes used by foreign governments, different international aid agencies, and non-governmental organizations vary dramatically and need to be seamlessly harmonized in order to effectively assist those stricken by the disaster." The lack of effective data management solutions can lead to a lack of coordination and chaos. The result can be delayed or ineffective response, the potential wastage of pledged support, imbalances in aid distribution, and a lack of transparency. This has been demonstrated both in disasters in areas with poor information and communication technology (ICT) infrastructure, e.g., the 2005 Kashmir/Pakistan earthquake, as well as during the 2005 hurricanes Katrina and Rita.

UMIACS faculty member Louiqa Raschid (who also holds an appointment in UM's Smith School of Business) and colleagues Craig Knoblock at the University of Southern California and Felix Naumann at the University of Potsdam have developed a research agenda for "On-the-fly Information Integration for Disaster Data Management."

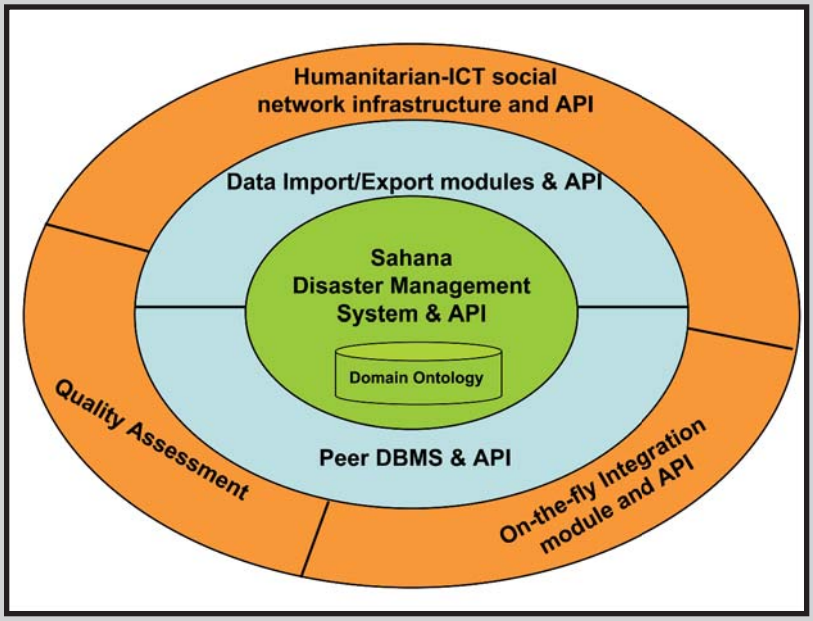
According to Raschid, "Applying ICT to disaster management is not a new concept. However, despite the tremendous value of ICT, there are very few disaster management systems that exist today and none are widely deployed or web-accessible. There is no single cohesive system that organizations such as the United Nations Disaster Assistance and Coordination (UNDAC) can routinely deploy so that ICT can be used effectively from one disaster to the next."

Raschid's research is based on the philosophy that disaster data management (DisDM) challenges have to be addressed within a flexible on-the-fly integration framework which effectively harnesses human expertise and human processing power. On-the-fly integration requires a flexible architecture to access and integrate sources. Initially, as sources are discovered, they need to be integrated in an ad hoc mode, since they will not be carefully modeled, and there may be little metadata. Subsequently, as the source is modeled in more detail, and as users access the source and assess its quality, it may be further integrated in a more principled engineered manner, with clear specifications of the integration objectives. According to Raschid, "Social networks are gaining widespread popularity on the web and we believe a social networking model could work well for a disaster management site." Some of the properties of disaster management are that the response typically involves many distributed and unrelated agencies and organizations, there are many organizations and individuals that have relevant information to share, and all of the agencies, organizations, and individuals are motivated to help in the disaster response. These factors make a social networking model particularly attractive to address the critical challenges of DisDM, which include finding, sharing, and evaluating relevant information.

Professor Louiqa Raschid

Raschid's involvement in applying ICT to disaster management goes back to the 2004 Asian tsunami. The Sahana project (http://sourceforge.net/projects/sahana/) was born as a Free Open Source Software (FOSS) application in Sri Lanka to address disaster management, collaboration and coordination in the aftermath of the tsunami. Sahana is distributed from SourceForge under the terms of the GNU LGPLicense. Raschid served as the Database Architect for the project and is a member of the Sahana Board. The Sahana project has gained widespread attention from developers and humanitarian consultants worldwide. It was recognized as SourceForge Project of the Month, June 2006, a 2006 Google Summer of Code project and was a Finalist for the 2006 Stockholm Challenge. The IBM Crisis Response Team has deployed Sahana in 2005 and 2006 together with the Philippines National Disaster Coordinating Council (http://ndcc.gov.ph/ndcc/).

Raschid is excited to be able to combine her research interests with a humanitarian FOSS application. As a native Sri Lankan, she is also proud to be involved in a project that showcases the technological skills and the commitment of young Sri Lankans.



Architecture for a FOSS disaster management system.

Ben Shneiderman: Invention to Application, or How Research Ideas become Commercial Products

Necessity is said to be the mother of invention, but it is probably also true that preparation is the father of invention. The need to understand hierarchies is apparent to many, but only a few might have been as prepared as UMIACS researcher Ben Shneiderman (Professor, Department of Computer Science) to invent a new visual interface to explore tree structures.

In 1990, when dealing with a crowded shared disk used by 14 people, he was propelled to find an effective visualization that would show the proportion of space used by each person's files. After trying many alternatives based on traditional node-link drawings, he shifted to using areas to show file sizes. But packing all the files into one rectangular screen without wasting space proved to be a challenge. Then in the faculty coffee lounge, he had a well-remembered AHA! experience. His excitement was intense, but it took him three days to work out the details and convince himself that it was a general solution.

The original paper was published in the prestigious *ACM Transactions on Graphics* in 1992. Graduate student Brian Johnson produced the first implementation, with many important improvements as part of his doctoral dissertation. Over the next few years, five implementations by students explored variations and produced a dozen papers that are still frequently referenced (See treemap history page http://www.cs.umd.edu/hcil/treemap-history/).

The temptation to stay with a successful research topic is great, but by

Treemap showing part of a hard drive with folders for recent years, with size of each rectangle proportional to file size in bytes. Color indicates file type, such as red for pdf, light red for ppt, black for zip and yellow for doc. This is a well-organized hierarchy but there is a duplication of two large green bmp files in the 1999 folder.

1997 Shneiderman thought that other problems were more interesting. He worked on photo libraries, government statistical data sets, menu design, and began his explorations of creativity support tools; meanwhile treemap innovations were being made by other researchers and commercial developers.

Enchanted by what he saw, he became a consultant for Smartmoney.com whose stock market treemap (MarketMap), implemented by Martin Wattenberg, became a sensation. Other research and commercial ventures showed still more possibilities for extending the treemap idea, so by 2000, Shneiderman was back into doing more treemap research. As part of the May 2001 Human-Computer Interaction Lab Annual Symposium, Shneiderman organized a day-long treemap workshop, bringing 20 researchers and commercial developers together. One of the presenters was Marc Smith of Microsoft Research who led a group that developed a new version of treemaps for understanding Usenet discussion group

continued on page 6...

UNIVERSITY OF MARYLAND

Institute for Advanced Computer Studies
2119 A.V. Williams Bldg.
University of Maryland
College Park, MD 20742-3311

Nonprofit Org.
U.S. Postage
PAID
ICM