

iIS Update

資訊科學簡訊

Newsletter of the Institute of Information Science
Academia Sinica, Taiwan



0011100 00

00001



000

011010

244IIS

0001010

IIS10100 10

IIS 101010 100

00010111

Institute of Information Science

03210.

00104

0000041

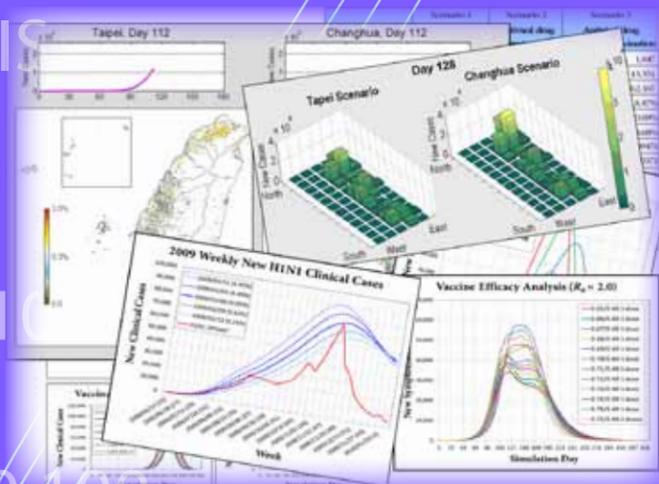
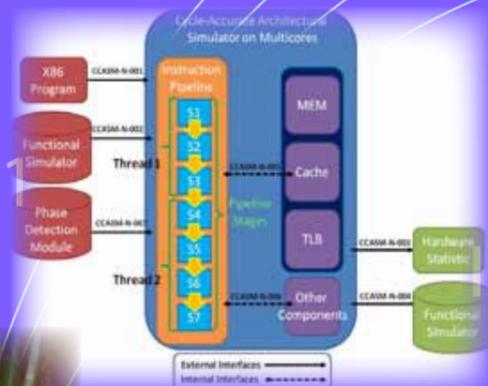


00104

00104

What's Inside

- 2 Message from the Director**
- 3 Honors and Awards**
- 4 Lab Profile**
Unlimiting the Computer:
Computer Systems Lab
- 6 Distinguished Lecture Series**
- 7 Project**
OSSF: Passionate about the
Spirit and Development of
Open Source Software
- 8 Activities**
- 10 Project**
Efficient Simulation of
the Spatial Transmission
Dynamics of Influenza
- 12 Spotlight**
Dr. Yeh hopes to uncover
greater knowledge from
various types of data ...
- 14 Great Ideas**
From the Netherlands to
Poland



Institute of Information Science

The Institute of Information Science (IIS) was established in 1982. We currently have 38 full time research faculty, 30 post-doctoral research fellows, and slightly more than 300 research associates and specialists. Our research is composed of eight specialized laboratories: Bioinformatics, Computer Systems, Information Processing and Discovery (iPAD), Multimedia Technology, Natural Language and Knowledge Processing, Network Systems and Services, Programming Languages and Formal Methods, and Computation Theory and Algorithms.

IIS is not a degree-granting academic institution, with the important exception of the international graduate program in bioinformatics, under the auspices of Academia Sinica's Taiwan International Graduate Program. This Ph.D. program was established in 2002 and has enrolled about 31 students over the last seven years.

Many of our research fellows hold joint faculty appointments at top universities in Taiwan. This allows our institution to play a very significant role in training and fostering advanced research talent in the IT industry and academia in Taiwan.

DIRECTOR:

Dr. Yew, Pen-Chung

DEPUTY DIRECTORS:

Dr. Ko, Ming-Tat

Dr. Chen, Keh-Jiann

GROUP COORDINATORS:

Dr. Sung, Ting-Yi

Bioinformatics Lab

Director Yew, Pen-Chung

Computer Systems Lab

Dr. Chen, Meng Chang

Information Processing and Discovery (iPAD)

Dr. Liao, Hong-Yuan Mark

Multimedia Technology Lab

Dr. Hsu, Wen-Lian

Natural Language and Knowledge Processing Lab

Dr. Ho, Jan-Ming

Network Systems and Services Lab

Dr. Chuang, Tyng-Ruey

Programming Languages and Formal Methods Lab

Dr. Ko, Ming-Tat

Computation Theory and Algorithms Lab



Message from the Director

Industries in the information technology (IT) sector have been the driving force of Taiwan's economic engine and will remain the main thrust of the country's continued economic growth for the foreseeable future. There is no industry today that could survive without using some kind of IT infrastructure or without the support of IT-related products. The IT revolution has been so fundamental and ubiquitous that it has become hard to imagine a world without it. And this revolution's changes will only accelerate as time goes on.

Academia Sinica's Institute of Information Science (IIS) is at the forefront of research in many IT-related fields. This issue of our newsletter highlights one of IIS's research laboratories, the Computer Systems Laboratory (CSL), whose main research areas include the latest multi-core systems, system virtualization, cloud computing, and system-design methodologies.

CSL's focus on multi-core systems covers the design of future generations of server-on-chip systems that support a many-core environment, in particular, cache memory hierarchy design. The system virtualization effort is concentrated on the core technology of dynamic binary translation, which could allow a guest binary code compiled for one instruction-set architecture (e.g., ARM's instruction set) to be run on a host platform with a different instruction-set architecture (e.g., Intel's x86). The effort targets multi-threaded applications running on multi-core platforms. Cloud computing today supports many large-scale applications; CSL is looking particularly at large-scale Internet applications such as online games. The design of database systems for easy and efficient data retrieval and effective file and I/O system support are the main research focuses. CSL is also researching multi-core design methodologies, in particular fast cycle-accurate system simulation techniques (critical for perfecting low-power designs) that could leverage multi-cores themselves, (i.e., use multi-cores to simulate multi-cores for fast simulation turnaround time).

Another major research project highlighted in this newsletter uses large-scale data, such as Taiwan's census data and international airport traffic analysis, to simulate the spread of infectious diseases with epidemic potential, such as the H1N1 flu. IIS's Information Processing and Discovery Laboratory (iPAD) has developed highly effective algorithms that speed up such simulations more than 1000-fold. The simulations can be used to help determine the impact of public health policies, such as inoculation of different age groups, on the spread of diseases when the supply of vaccines is limited due to the lag time in vaccine production. This technology has already proven itself highly valuable, including during the H1N1 flu pandemic of 2009.

The iIS Newsletter strives to be both informative and entertaining. We hope you will find Dr. Shin-Cheng Mu's Great Ideas interesting. We encourage our readers to tell us what you would like to see in this newsletter series. Please send your comments and suggestions to iis_newsletter@iis.sinica.edu.tw.

Honors and Awards

- ◆ **Dr. Tsan-sheng Hsu's** computer game research team (which includes Bo-Nian Chen, Bing-Jie Shen, Meng-Tsung Tsai, and Hung-Jui Chang) won a total of 1 gold, 1 silver, and 2 bronze medals in the 2nd TAAI Cup and 15th Computer Olympiad.
- ◆ **Dr. Shen-Wei Chen** received the 2010 Outstanding Young Electrical Engineer Award from the Chinese Institute of Electrical Engineering.
Dr. Chen was awarded the prize for his outstanding research contribution to network and multimedia quality of experience research and online game research. His research group is one of the few dedicated to issues related to the design and operation of online games. Their research areas cover network performance, user satisfaction, cheating, user-behavior analysis, and computer-human interaction in online games. Dr. Chen actively participates in academia-industry collaborations, with the long-term goal to make changes in the digital entertainment industry by bringing together efforts from both academia and industry.
- ◆ **Dr. Chun-Nan Hsu's** AIIA Lab was ranked first in the BioCreative III Challenge evaluation of biological text mining. The task of gene normalization is to convert a gene mention in a text to a standard gene ID. The task is challenging due to the absence of a comprehensive dictionary and high ambiguity among genes of different species. When nine different methods were used to rank competing systems, the AIIA's system was ranked at the top by one method and consistently in second or third place by all the others.

2011 Distinguished Lecture Series

January	Dr. Herbert Edelsbrunner	Institute of Science and Technology, Austria	Algorithms, Discrete and Computational Geometry, Computational Topology, Applications to Biology
January	Dr. Jean Ponce	Ecole Normale Supérieure	Computer Vision, Computer Graphics, Machine Learning, Robotics
February	Dr. John Rushby	SRI International	Design and Assurance of "Critical Systems"
March	Dr. Moshe Vardi	Rice University	Applications of Logic to Computer Science
April	Dr. Stephane Mallat	Ecole Polytechnique	Applied Mathematics for Signal and Information Processing, as well as in Computer Vision
May	Dr. Qiang Yang	Hong Kong University of Science and Technology	Artificial Intelligence, Data Mining and Machine Learning, Activity Recognition
June	Dr. Wen-Mei Hwu	University of Illinois at Urbana-Champaign	Computer Architecture, Compilers

• Information is subject to change. Please check our website for the latest schedule •

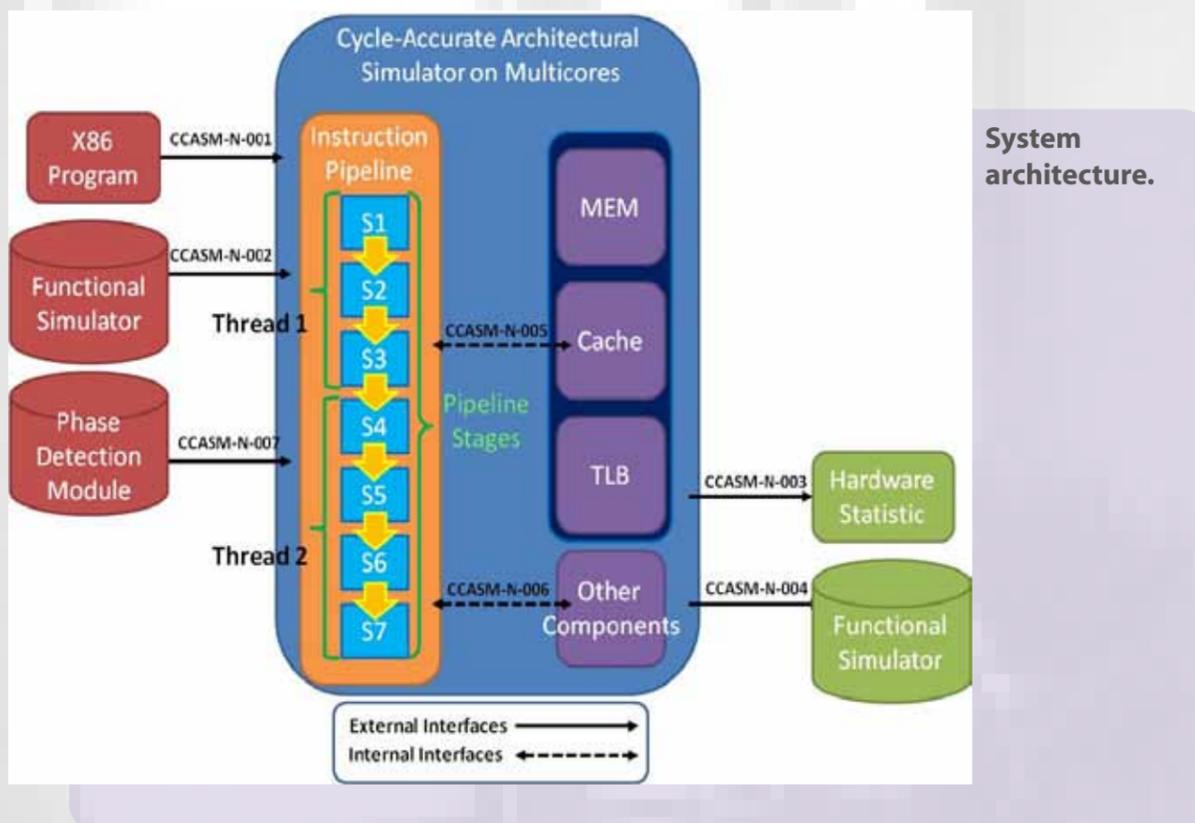
TIGP Bioinformatics Program Distinguished Lecture Series



As part of the 2010 TIGP Distinguished Lecture Series, Dr. Wen-Lian Hsu addressed the topic "Intelligence: Machine vs. Human" on November 12. In this lecture, Dr. Hsu shared his decades of research experience in the development of artificial intelligence and discussed how it has helped to broaden the spectrum of research as well as to improve the quality of research in various disciplines.

TIGP Bioinformatics Distinguished Lecture Series: Left, Vice President Chao-Han Liu. Right, Dr. Wen-Lian Hsu

Unlimiting the Computer: Computer Systems Lab



The Computer Systems Lab was established in 2009. Its primary research areas include multicore systems, virtualization, and system software for cloud computing and related applications.

1. System Design and Design Support for Multicore Systems

This research effort focuses on the architecture and system support needed for various new system features. Designing such multicore systems requires extensive simulations at the system level before hardware design can begin. In order to shorten critical time-to-market, software development needs to proceed on simulators at the same time hardware is being designed and implemented. Hence, speeding up these simulations will facilitate faster design time and produce better system designs.

Our approach is to use existing mul-

ticore platforms to simulate new multicore designs by exploiting their inherent thread-level parallelism. To speed up such simulations, various existing techniques for single cores are extended to multicores. For example, we exploit similar program behavior in program-execution phases. We also integrate fast functional simulations with detailed cycle-accurate simulations in order to both shorten simulation times and to obtain accurate simulation results.

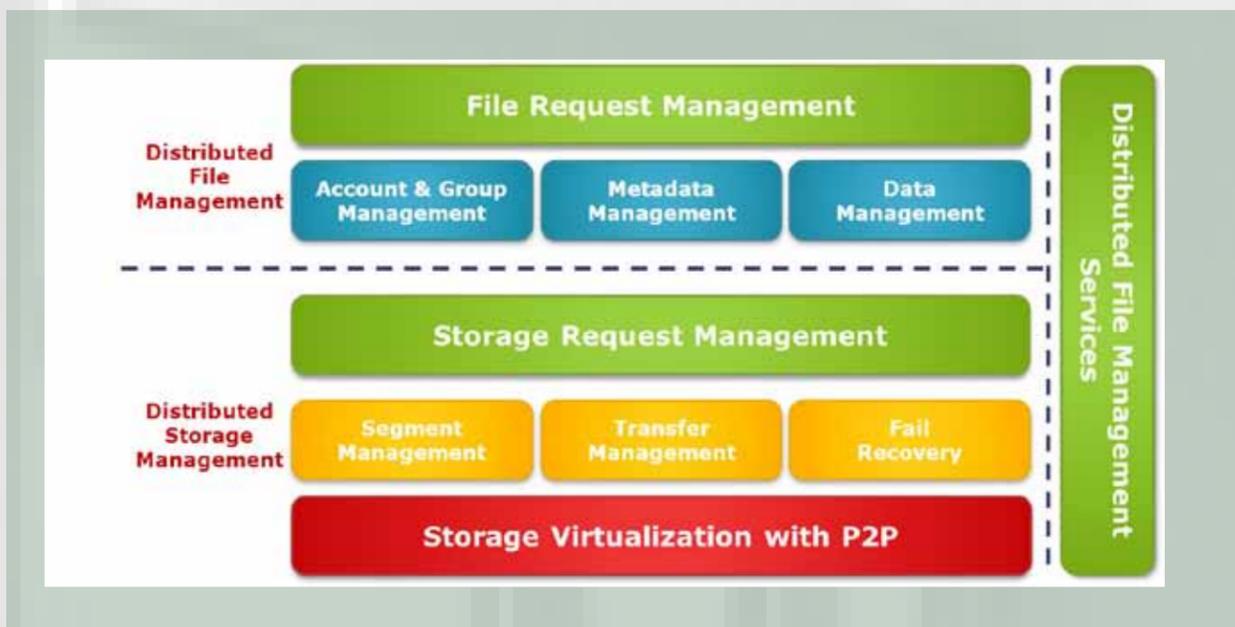
2. Parallel Programming Support for Multicores

Software applications are the ultimate determining factor for the success of multicore technologies. However, parallel applications are not only notoriously difficult to program but also hard to debug because of the indeterministic interaction between threads on such platforms. Thus, tools to support performance tuning and efficient debugging on multicores are very useful to programmers.

Various domain-specific parallel programming models have been proposed to help application programmers migrating their existing applications or writing new code for new multicore platforms. Recent successful examples, such as CUDA from Nvidia and MapReduce from Google, have been extensively used in computer graphics and search engines. They are also being explored in other application domains.

Current research efforts include ex-

Storage virtualization with P2P.



ploring new techniques for more efficient and effective debugging support for parallel applications. This would allow application programmers to contribute domain knowledge that could guide compilers in generating high-performance parallel codes on multicores, as well as on systems with general-purpose graphic processing units (GPGPUs).

3. System Support for Virtualization

System virtualization is a very important technology for multicores and cloud computing. It allows applications running on such systems to be agnostic about the underlying platforms. This research effort focuses on the core technologies, in particular, dynamic compilation techniques for binary translation and binary optimization targeting multi-core systems.

The research team is exploring new ways of extending compilation techniques for single-core to multicores. One of the goals is to take binary codes from multiple instruction-set architectures and translate them to another set of instruction-set architectures on multicores. All of such binary manipulations are under one unified dynamic compilation framework. Furthermore, this occurs at the binary-code level, instead of using

high-level language virtual machines such as Java and C#. It affords much higher transparency and portability for existing binary codes.

4. Resource-Efficient Cloud Middleware for Highly Interactive Application Services

Supporting millions of avatars simultaneously in a networked virtual environment (NVE) is not feasible for a physical server; therefore, the NVE needs to be partitioned over multiple servers. However, the partitioning of an NVE is a difficult problem because 1) the avatars in an NVE are not uniformly distributed; and 2) the distribution of the avatars may change over time.

This research focuses on resource-efficient cloud middleware for highly interactive application services, with the goals of supporting massive multi-player NVEs, improving efficiency and power usage on servers, and guaranteeing real-time interactivity. This demands innovative solutions to challenging issues, such as performance assurance and fairness in virtualization, power/resource management, workload prediction, fast live migration, failure recovery, in-memory database design, and dynamic data partitioning/replication/localization.

5. Adaptive Data and Storage Management in Distributed File Systems for Clouds

Cloud computing is a new and promising paradigm in which dynamically scalable and often virtualized resources are provided as a service over the Internet. It leverages networks and integrates distributed computing and storage resources as a virtual platform on which users can not only access provided application services, but also execute large-scale distributed applications that require significant amounts of computing and storage resources. The Google File System (GFS) is a pioneer in this area.

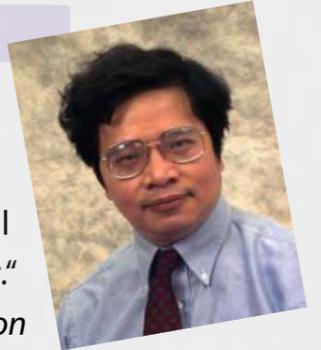
This research effort aims to explore adaptive data and storage management for different workloads and technological environments, as both performance and efficiency will improve if data and storage-management strategies can adapt to these different workloads and environments. This demands innovative solutions to challenging issues, such as workload profiling and clustering, inter-file association, environmental benchmarking, data and storage management strategies for different workloads and environments, online monitoring, and workload classification, as well as the selection of management strategies.

Distinguished Lecture Series



"In order to improve learning effectiveness, a potential paradigm shift is to design the objective function to match the expected performance of the desired decision-making process so that feedback can be collected with the current model to improve parameter learning in an interactive manner."

—Dr. Chin-Hui Lee. August 2, 2010. *Learning to Match Decision Making: A Decision-Feedback Learning Paradigm*



"I will present a research plan toward fine-grained semantic information access via NLP and related research challenges."

—Dr. Jun'ichi Tsujii. September 6, 2010. *Semantics-based Information Access to Large Scale Text Bases*

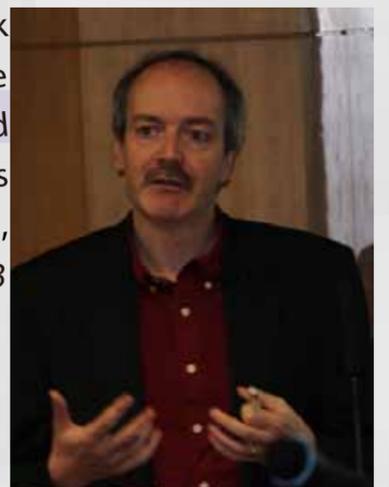


"This talk will describe our attempt to build a never-ending language learner, NELL, that runs 24 hours per day, forever, and that each day has two goals: (1) extract more structured information from the Web to populate its growing knowledge base, and (2) learn to read better than yesterday, by using previously acquired knowledge to better constrain its subsequent learning."

—Dr. Tom Mitchell. November 17, 2010. *Never-Ending Learning*



"Cryptographic hash functions are an essential building block for security applications.... In spite of this central role in applications, the amount of theoretical research and cryptanalysis invested in cryptographic hash functions was rather limited." —Dr. Bart Preneel. December 1, 2010. *Cryptographic Hash Functions and the SHA-3 Competition*



OSSF: Passionate about the Spirit and Development of Open Source Software



OSSF members, "Work hard, play hard" is their motto. (Middle) Principal investigator: Dr. Jane Win-Shih Liu.

The Open Source Software Foundry (OSSF) was founded in the summer of 2003 by the Ministry of Economic Affairs' Industrial Development Bureau to help the development of open source software (OSS) and to expand the knowledge and use of such software in Taiwan. What began in a small office in Academia Sinica's Institute of Information Science (IIS) has grown to a 15-member team in

the Research Center of Information Technology Innovation (CITI). The operational budget now comes jointly from National Science Council and CITI.

The OSSF family consists of people who are passionate about the spirit and development of open source software. To achieve their annual goals, they have specific jobs to do; these jobs are divided into three categories: Technical Development, Outreach and Promotion, and License and Legal Consultation. All OSSF members are under the supervision of four principal investigators: Dr. Lee Der-Tsai, Dr. Liu Win-Shih, Dr. Chuang Tyng-Ruey, and Dr. Hsiao Ching-Teng.

The OSSF provides several tools and services for all OSS developers and users via the OpenFoundry website (www.openfoundry.org), a portal site for experience sharing and a platform for relevant OSS components and human resource references. The OSSF also provides five major services: an OSS license and legal

(cont'd on page 15)



Open source software technology is continuing to advance.

Activities

Visitors



Dr. Frank D. Dennis, founder of Taitung Christian Hospital, and his wife, Ms. Sally Dennis, visit on September 28, 2010. From left: Director Yew, Dr. Frank Dennis, Dr. Ting-Yi Sung, Ms. Sally Dennis.



A delegation from the Polish Ministry of Science and Higher Education visits on November 9, 2010. From left: Director Yew, Dr. Der-Tsai Lee, Dr. Maria E. Orłowska, Dr. Leszek Grabarczyk, Mr. Yeh-Shin Chu.

Dr. David Atkinson and Dr. Peter Friedland of Asian Office of Aerospace Research and Development in Air Force Office of Scientific Research visits on October 25, 2010. Right Dr. David Atkinson. Below: Dr. Meng Chang Chen, Director Yew, Dr. Peter Friedland, Dr. Wen-Lian Hsu.



Forum on Computer Games held on September 4, 2010, at the Institute of Information Science, Academia Sinica, Taipei, Taiwan. Chairman: Prof. Tsan-sheng Hsu (middle of the first line)



A team from the Ministry of National Defense visits to inspect those who are part of the Defense Industry Reserve Duty system.



Research assistants demonstrate their projects to spectators.

2010 Open House

9:00 a.m. – 4:00 p.m., Saturday, October 23

After enduring days of dreary weather and heavy rain, we were delighted to be greeted on Saturday morning by sunshine and a cool breeze. We walked through our exhibition areas to make sure that everything was ready, and then opened our doors to a crowd of people who were already waiting to enter.

IIS's 2010 Open House was a great success. Thank you to those who were there to make it happen. Without all your hard work, we would not have made it. Great job, everyone!



People line up in front of the IIS 2010 Open House billboard.

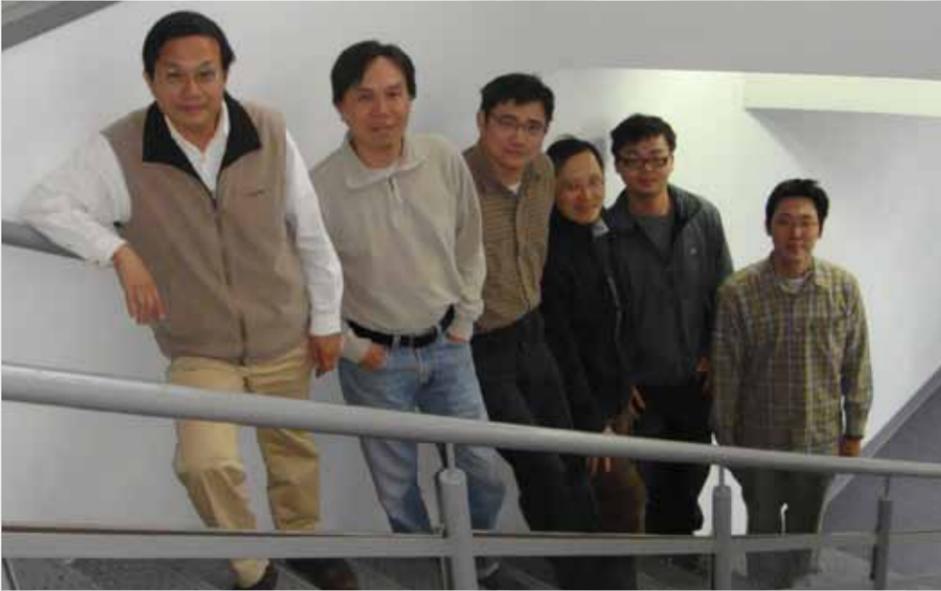


Visitors examine one laboratory's display.



Balloons add to the festive atmosphere.

Efficient Simulation of the Spatial Transmission Dynamics of Influenza

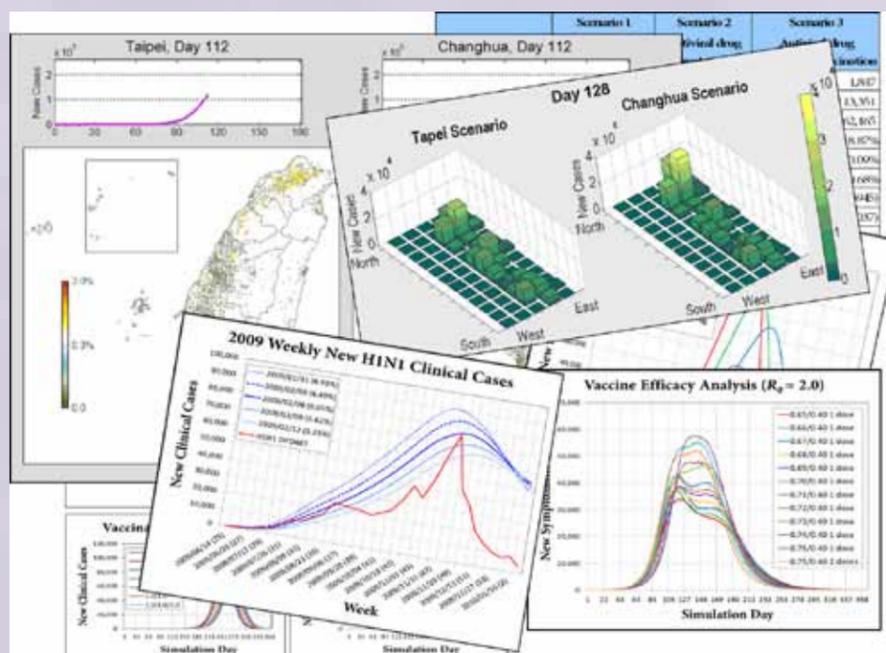


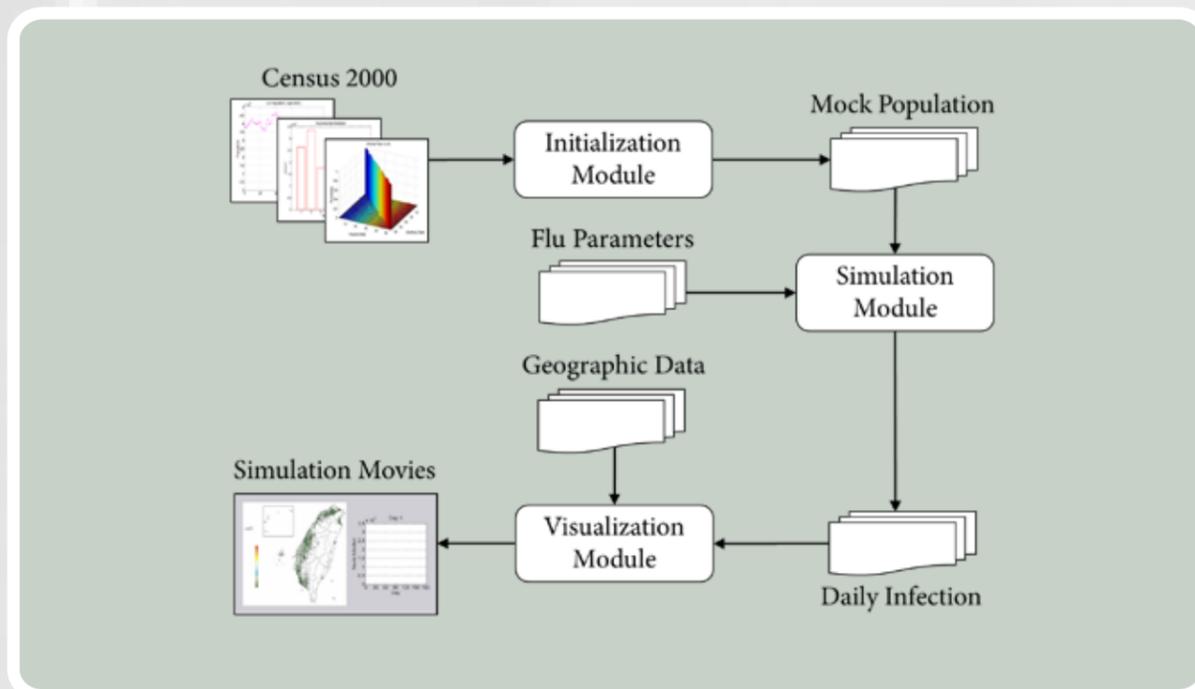
Group members (left to right): Dr. Tsan-sheng Hsu, Mr. Tsurng-Chen Chen, Dr. Da-Wei Wang, Dr. Churn-Jung Liao, Mr. Meng-Tsung Tsai and Mr. Hung-Jui Chang.

Computer simulations have been used for decades to help understand the patterns by which diseases spread. Certain diseases, such as pandemic influenza, follow spatial patterns. Data from the 2009 H1N1 pandemic (H1N1pdm) suggest that previous studies overestimated the within-country rate of spatial spread of pandemic influenza. As large spatially resolved data sets are constructed, the need for efficient simulation code with which to investi-

gate the spatial patterns of pandemics becomes clear. Here, the new system presents a significant improvement to the efficiency of the individual based stochastic disease simulation framework commonly used in previous studies. The new system quantifies the efficiency of the revised algorithm and presents an alternative parameterization of the model in terms of the basic reproductive number. The system applies the model to the population of Taiwan and demonstrates how the location of the initial seed can influence spatial incidence profiles and the overall spread of the epidemic. Differences in incidence are driven by the relative connectivity of alternate seed locations. The ability to perform efficient simulations allows us to run a batch of simulations and take account of their average in real time. The averaged data are stable and can be used to differentiate spreading patterns not readily apparent when conducting only a few runs.

This research is being conducted by





members of the Institute of Information Science's Laboratory of Massive Data Computing and Management, led by Dr. Tsan-sheng Hsu, Dr. Churn-Jung Liao, and Dr. Da-Wei Wang, with joint efforts from experts in Taiwan's Centers for Disease Control, Imperial College London, and National Taiwan University.

The real challenge of this study is to incorporate census data for each citizen in Taiwan with spatial data representing the traveling patterns and geographic information such as household, school and commuting for possible contacts. This represents an enormous amount of data, the computation of which would normally require a super computer rather than a desktop server. But breakthroughs in fundamental computing algorithms have made our system several hundred times faster than previous similar systems. In designing the model, we have also considered the issue of data privacy, which is another longtime concern of the Massive Data Computing and Management Laboratory. Our system can output the general behavior during the simulation in a way that no individually identifiable information can be deduced from it.

Preliminary results were published in PLoS ONE [1], a journal that covers primary research from disciplines within science and medicine, as well as presented at a well-known international conference

on theoretical computer science [2]. That this interdisciplinary research has been recognized by experts in both computer science and public health testifies to its success. Using a fundamental study in computer science to build a simulation system that has practical applications is also a valuable experience for researchers.

Last year Taiwan's CDC developed a system based on our theoretical studies to help prepare for a possible H1N1 pandemic flu outbreak. Using the fast simulation modules, the CDC was able to simulate the possible outcomes of deploying certain intervention policies. The results were an important reference in implementing Taiwan's intervention strategies. For example, the study suggested that implementing tight airport quarantine policies could delay the peak date of H1N1 infection by up to 22 days, which has proven to be a very good estimate.

This study can be used to solve problems in many other domains that require a model for interactions between people. The fundamental algorithms can be used to solve other problems of data-intensive computing. Further research problems for this study is to incorporate more personal information into the model to make it closer to the real world.

1. M.-T. Tsai, T.-C. Chern, J.-H. Chung, C.-

W. Hsueh, H.-S. Kuo, C.-J. Liao, S. Riley, B.-J. Shen, D.-W. Wang, C.-H. Shen and T.-S. Hsu, "Efficient Simulation of the Spatial Transmission Dynamics of Influenza", PLoS ONE, November 4 2010.

2. M.-T. Tsai, D.-W. Wang, C.-J. Liao and T.-S. Hsu, "Heterogeneous Subset Sampling", Proceedings of the 16th International Computing and Combinatorics Conference (COCOON), Springer-Verlag LNCS# 6196, pp. 500-509, 2010.

Share your photos in iPicBox.tw.

A Knowledge and Information Explorer: Design Better Mining Algorithms for Data with Different Characteristics



Mi-Yen Yeh at VLDB 2008, Auckland, New Zealand.

Dr. Mi-Yen Yeh is an assistant research fellow at the Institute of Information Science and holds a joint appointment at Academia Sinica's Research Center for Information Technology Innovation (CITI). She received her B.S. and Ph.D. degrees, both in electrical engineering, from National Taiwan University in 2002 and 2009, respectively. Prior to joining IIS, she was a postdoctoral fellow at CITI from February to September 2009.

Dr. Yeh's research interests lie in the area of data mining and databases. She focuses especially on mining data streams, including stream summarization, clustering, and similarity query processing. She has published two papers in *IEEE Transactions on Knowledge and Data Engineering*, the top journal in the field of database/data mining.

More importantly, the results of one of her research projects were presented at the 34th International Conference on Very Large Data Bases (VLDB 2008), the most prestigious conference in the databases field. VLDB accepts only about 16% of papers submitted and had not accepted another paper by a Taiwan student

since 2004.

Supported by the NSC Graduate Students Study Abroad Program, Dr. Yeh spent about one year, in 2007 and 2008, at IBM T.J. Watson Research Center, a world-renowned industrial research organization. Through interaction with many of the outstanding researchers and professionals there, Dr. Yeh learned new research ideas, acquired new problem-solving skills, and was further inspired to pursue a career in research. Their passion and dedication to research was infectious!

Dr. Yeh hopes to uncover greater knowledge from various types of data such as vehicle trajectories and social networks.

Currently, Dr. Yeh has the following research directions.

1. Data stream mining and management

Advanced database technology and automated data collection tools gather vast amounts of data. Data generated orderly and continuously and taking the form of a stream are referred to as a data stream. Processing data streams is increasingly important, as more and more emerging applications are required to handle this kind of data. Examples in-



clude data analysis in sensor networks and program trading in financial markets. Moreover, in many cases more than one data stream needs to be analyzed simultaneously. Therefore, to understand and discover the interesting patterns within data streams, one needs to design algorithms that are limited-pass, real-time, and have bounded memory usage. Dr. Yeh hopes to extend her previous research work as well as to keep tracking other related issues in the data stream environment.

2. Trajectory data mining

With the common use of tracking and positioning systems such as GPS devices and the advance of mobile communication technologies, a huge amount of trajectory data can be collected almost anytime, anywhere. The trajectories are usually recorded with both positions and timestamps denoting the displacements and the speeds of objects being tracked. For example, the courses of different types of vehicles can be recorded with the equipped GPS devices, and the movement paths of migrating animals can be recorded by attached transmitters. By exploring the special moving behaviors of monitored vehicles in Taipei, Dr. Yeh hopes to design better mining algorithms that consider both the spatial and temporal characteristics of the trajectories.

3. Social network analysis

Social network analysis has attracted more and more attention from the data mining community in recent years. By modeling the social network as a graph structure, where a node is an individual and an edge represents the relationship between individuals, many studies have addressed graph mining techniques to discover interesting knowledge from such networks. Given advances in communication technologies and the explosion of social web applications such as Facebook and Twitter, the scale of the generated network data is usually very large — so large that attempting to store all the data from the enormous network before working to discover the characteristics of these social networks appears prohibitively difficult. Therefore,

Dr. Yeh hopes to develop an efficient and systematic approach to gathering data of an appropriate size while preserving as much as possible the properties of the original network, especially heterogeneous networks.

To learn more about Dr. Yeh, please visit her website at www.iis.sinica.edu.tw/~miyen.

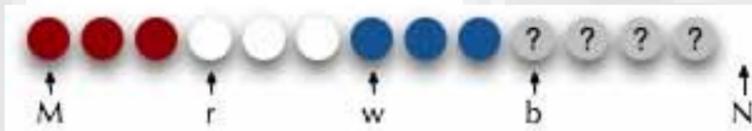
The 2010 Conference on Technologies and Applications of Artificial Intelligence

was held in Hsinchu, Taiwan, November 18–20.

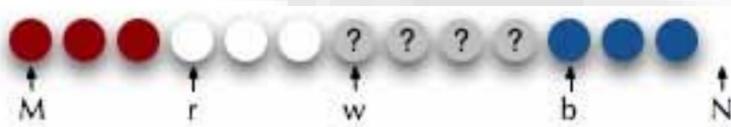
This annual gathering, now in its 15th year, is one of the most important regular academic meetings in Taiwan on artificial intelligence. It is sponsored by the Taiwanese Association for AI.



From the Netherlands to Poland



One possibility is to allow an “unknown” section on the side. (Fig. 1)



The “unknown” section between two colors, say white and blue. (Fig. 2)



The invariant they used. (Fig. 3)

By Shin-Cheng Mu

Given an array of elements, each having one of three colors: red, white, and blue. How can the elements be swapped such that the array is sorted in the order red, white, and blue?



This “Dutch national flag” problem was proposed by Feijen and Dijkstra (Why not French or Yugoslavian? Apparently both Feijen and Dijkstra were Dutch!). It is well known that quicksort performs badly for arrays that are almost sorted or contain repeated elements. Let red, white, and blue represent elements that are, respectively, less than, equal to, and greater than the pivot. Solving the Dutch national flag problem leads to a three-way partition that handles repeated elements better.

To construct the program, we know that we definitely need a loop. And

whenever there is a loop we shall be constructing a loop invariant. One possibility is to allow an “unknown” section on the side (see Fig. 1).

Or we may leave the “unknown” section between two colors, say white and blue (see Fig. 2).

One of the invariants leads to a better algorithm. Which will it be?

Until the mid 1990s, conventional wisdom was that the three-way partition needs too many swaps and is not worth doing for quicksort. In the 1990s, most of the quicksort routines were derived from the one written by Lee McMahon for Version 7 Unix, which had been around for 20 years. The common hack was to switch to either bucket sort or radix sort when the array gets small. Still, Bentley and McIlroy noticed that, for all quicksort routines they could find, there was always an easy way to generate an input that had to be processed in quadratic time. So they set out to build a more robust quicksort, and studied the Dutch national flag problem again.

The excessive number of swaps in the traditional solution, they claimed, was

due to stepwise swapping the white elements to the middle. Instead, they recommended leaving the white elements on the two sides, sorting the array in the order “white, red, blue, white,” and finally swap the white elements to the middle in a separate pass (Fig. 3).

And it worked very well! The performance was significantly improved. The cost of the separate pass swapping white elements back can be amortized, since once they are in place, they need not be processed in the recursive call. Sedgwick and Bentley proved that such a quicksort is entropy-optimal, and it is in use in many Java and C libraries today.



However, the result is no longer the Dutch flag. There is not even a national flag with four stripes. But by swapping the colors a flag with four stripes (in blue, white, red, and blue) can be found that matches: that of Rypin, Poland, a town with a population of about 16,000.

References

Roland Backhouse. Program Construction: Calculating Implementations from Specifications. John Wiley & Sons, 2003.

Anne Kaldewaij, Programming: The Derivation of Algorithms, Prentice-Hall, 1990.

Jon L. Bentley, M. Douglas McIlroy. Engineering a sort function. *Software—Practice & Experience* 23(11), pp. 1249–1265. Nov. 1993.

Shin-Cheng Mu. The Dutch National Flag Problem. www.iis.sinica.edu.tw/~scm/ncs/2010/10/dutch-national-flag-problem/.

OSSF (story continued from page 7)

advisory service, the OSS biweekly newsletter, the OSS-related resource catalog, OSS learning materials, and the hosting of various OSS-related activities. By making these tools and services easily accessible, the OSSF hopes to help and encourage Taiwanese developers who are interested in the OSS field to develop new open source software.

But those in the OSSF aren't a bunch of geeks. Their motto is "Work hard, play hard!" So although they are intensely dedicated to their work and sometimes burn the midnight oil (the OSSF snack bar is open every day and always has hot

coffee ready), they don't forget to relax and have fun. Couch Potatoes? No! Jogging, dart games, and many other physical activities and competitions help project members relax, burn calories, and refocus their attention in a way that can spur creative thinking.

Please feel free to visit them in room 402 of the new IIS building. They are happy to share their experiences and knowledge about OSS development with one and all!

OSSF Project website: www.openfoundry.org/

Tel: +886-2-2788-3799 ext. 1469, 1477, 1478.

Workshop**The Rule of the GPL and Its Compliance Engineering - 20-Lecture Series IV**

Date: December 2, 2010

Category: Law and Licenses

Location: Conference Room 2, Humanities & Social Sciences Building, Academia Sinica

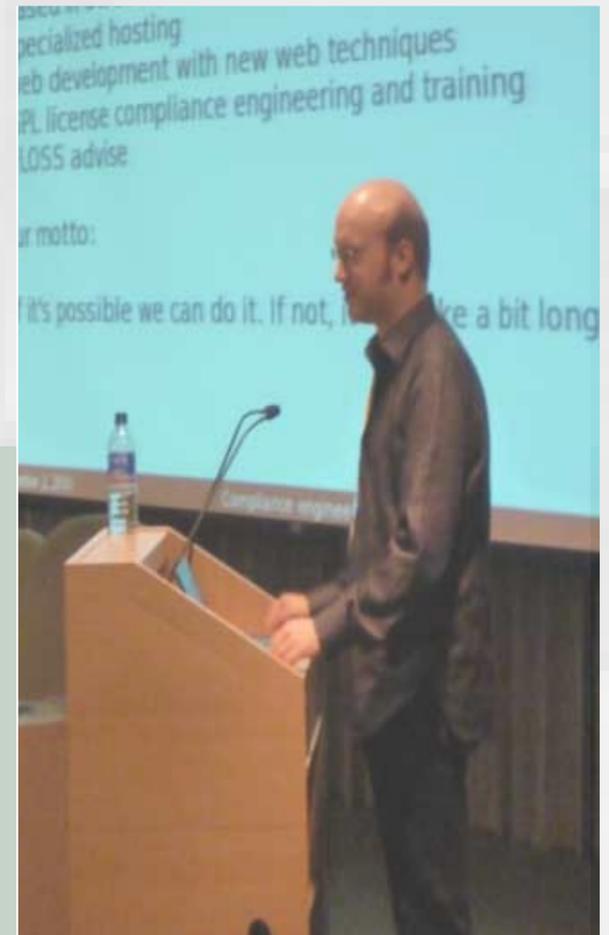
tel: 2788-3799 ext. 1474 or 0953-366-676

OSSF (Open Source Software Foundry) invited the core members of GPL-Violations.org, Harald Welte and Armijn Hemel, to be the distinguished speakers of this workshop. GPL-Violations.org is an organization focusing on GPL violation cases. Mr. Welte is the founder of the organization and the copyright holder of several well-known GPLed compo-



Mr. Harald Welte explains the content of the GPL and related lawsuits.

nents; in the past few years he has pursued several successful GPL-violation lawsuits in Europe. Mr. Hemel has for many years been in charge of his organization's



Mr. Armijn Hemel explains technical inspections.

technical inspection of GPL violations. Both of them have intensive knowledge of GPL and related technical skills.

For more information, please visit their website at www.openfoundry.org/en/workshop/details/115.

The Workshop on International Standardization of Web and Han Character Technology

was held on October 7. Six keynote speakers shared their experience in Unicode. The participants included Dr. Murata Makoto from the W3C XML Working Group and Yamamoto Taro from Adobe.



The Institute of Information Science (IIS) at Academia Sinica, Taiwan, ROC, invites all qualified candidates to apply for the positions of junior and senior research fellows of all ranks (equivalent to the ranks of tenure-track assistant, associate, and full professors in a regular academic department without teaching responsibility) in all areas of Computer Science. In particular, candidates in the areas of computer systems, machine learning, and natural language processing are strongly encouraged to apply.

Academia Sinica is a national academic research institution in Taiwan that conducts research on a broad spectrum of subjects in science and humanities. IIS is committed to high-quality research in computer and information science and engineering. In addition to research funding supported by Academia Sinica, external funding through government agencies and industry-sponsored institutions is also available.

Full-time research fellows are free to set their own research directions. IIS currently has about 40 full-time research fellows and close to 300 full-time post-doctoral fellows and research assistants. The areas of their current research include Systems Technology, Bioinformatics, Multimedia, Natural Language and Knowledge Processing, Network and Theoretical Computer Science, and Parallel Processing.

All candidates should have a doctoral degree in computer science or closely related fields, with a strong research and publication record. Senior candidates must demonstrate strong leadership and have an international reputation evidenced by publications, patents, industrial experiences, or other academic and scholarly achievements. Salary is commensurate with qualifications.

All candidates should send a detailed curriculum vitae and at least three letters of recommendation to

Director

Institute of Information Science
Academia Sinica
Nangang 115, Taipei, Taiwan
Tel: +886-2-2788-3799 ext. 1729
Fax: +886-2-2782-4814
E-mail: recruit@iis.sinica.edu.tw

Fluency in Chinese is an advantage, but not required. For additional information about IIS, please visit www.iis.sinica.edu.tw.



Institute of Information Science, Academia Sinica

128 Academia Road, Section 2, Nangang, 115, Taipei, Taiwan
Tel: 02-2788-3799
www.iis.sinica.edu.tw

Publisher: Director Pen-Chung Yew

Editors: Pei-Chi Wang and Huey-Chyi Chris Tseng