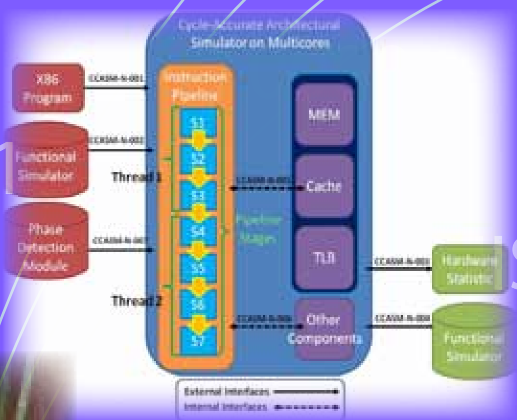
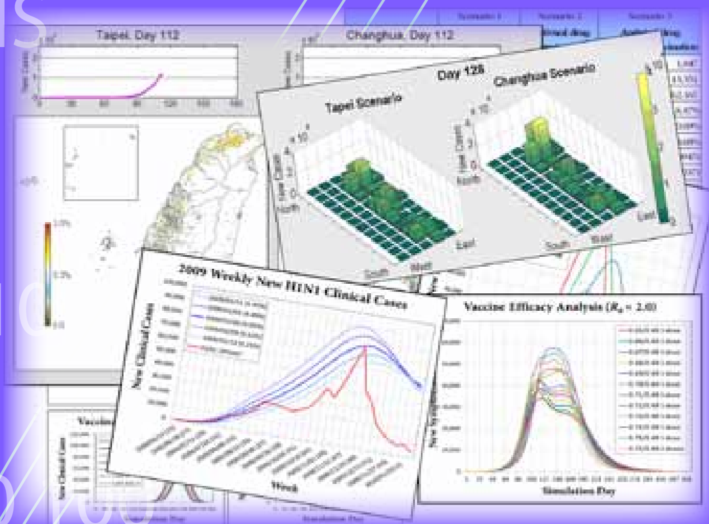
2011 春夏版

iIS Update

資訊科學簡訊

Newsletter of the Institute of Information Science
Academia Sinica, Taiwan





焦點報導

Institute of Information Science

資訊科學研究所於 1977 年開始設立籌備處，歷經五年籌備，於 1982 年 9 月正式成立研究所，是中央研究院數理組十個單位之一。本所的研究重點分為八大方向，包含生物資訊實驗室、電腦系統實驗室、資料處理與探勘實驗室、多媒體技術實驗室、語言與知識處理實驗室、網路系統與服務實驗室、程式語言與形式方法實驗室、計算理論與演算法處理。

本所並於 2002 年配合中研院國際研究生院 (TIGP)，開始招收生物資訊學程博士班，致力延攬及培訓此一新領域之國際人材，七年來已招募 31 名學生，對生物資訊的研究有很大的助益。

所內資深研究人員皆指導或共同指導來自全國各地頂尖大學之碩、博士生，並與這些大學保持緊密的研究關係及合作計劃，與國內外資訊業界亦有極為良好之互動關係。

所長：游本中

副所長：高明達
陳克健

研究群主持人：
宋定懿教授
生物資訊實驗室
游本中教授
電腦系統實驗室
陳孟彰教授
資料處理與探勘實驗室
廖弘源教授
多媒體技術實驗室
許聞廉教授
語言與知識處理實驗室
何建明教授
網路系統與服務實驗室
莊庭瑞教授
程式語言與形式方法實驗室
徐讚昇教授
計算理論與演算法處理實驗室

所長的話

資訊業一直是驅動台灣經濟的主導力量，在可預見的未來它仍將持續推動台灣的經濟成長。時至今日，沒有任何產業能夠不使用任何資訊架構或資訊產品而存活。資訊革命造就了人類基本上及全面性的改變，無法想像缺少了它世界將如何運轉。因此它也加速了時代的演進。

中研院資訊所致力於資訊相關領域的前瞻性研究。這一期我們特別介紹了電腦系統實驗室 (CSL)。其研究包括了最先進的多核心系統、系統虛擬化支援、雲端計算、系統設計方法。其多核心系統之研究重點包含支援未來眾核心 (many-core) 之伺服器單晶片系統，尤其是運用在高速快取記憶體層次結構設計，此一研究預計可被運用於執行於多核心平台上之多執行緒應用軟體。系統虛擬化之研究重點在於，動態編譯技術如二進制碼轉譯核心技術，使二進制程式碼動態編譯一個指令架構（如 ARM 的指令集），並且能在多個不同的指令集架構間轉換（比如英特爾的 x86）。現今，雲端計算支援許多大型的應用。當前研究主要應用於大規模高度互動式網路應用程式及服務之系統支援，例如線上遊戲。此研究目標在設計一資料庫系統，以提供簡單而快速資料存取及高效率之檔案和輸入輸出系統。此外，電腦系統實驗室 (CSL) 亦研究最先進的多核心設計方法，尤其是利用多核心技術，以快速進行周期精確之系統模擬，如運用多核心來模擬多核心以大幅縮短模擬所需時間。此一周期精確之系統模擬對於低功率系統設計有著舉足輕重的影響。

本期同時介紹了另一個由巨量資料運算實驗室所執行的重點研究計劃。他們運用大量的數據，例如台灣的國際機場普查數據和流量分析，模擬流行性傳染病散播的機率，如 H1N1 流感。該團隊發展高效率的演算法加速了模擬計算之時間超過一千倍。它同時模擬運用於研究流感防治政策之影響，疫苗因延遲生產而導致供應量有限，針對不同年齡層接種疫流感疫苗時，會有何成效。在 2009 年 H1N1 流感大流行時，已被證明是非常有用。

資訊科學簡訊兼具了知識性及娛樂性。我們希望在穆信成教授的創意發想 " 從荷蘭到波蘭 " 一文中，讓你們得到一些有趣的點子。我們更鼓勵讀者積極投入這一公開的園地，提供您們的意見與建議，或您希望在簡訊中閱讀到什麼內容，歡迎讀者投書至 iis_newsletter@iis.sinica.edu.tw。

◆ **徐讚昇研究員指導之電腦對局研究團隊（陳柏年，沈秉杰，蔡孟宗及張紘睿）在 2nd TAAI Cup 及 15th Computer Olympiad 獲得一金一銀二銅佳績。**

◆ **陳昇瑋助研究員榮獲 2010 年中國電機工程學會優秀青年電機工程師獎**

陳昇瑋博士所領導的團隊為全球少數專注於線上遊戲維運議題的研究團隊，從遊戲的網路效能、使用者滿意度出發，研究作弊行為及外掛偵測，到近期的使用者行為分析及人機互動發展，他積極地長期參與大型整合計畫與進行產學合作，將其所擁有的專業能力與豐富的實務經驗，貢獻給產業界。他期待能促進學術研究與產界實務的完美結合，為國內數位內容產業帶來新的視角及契機。

◆ **所研究員許鈞南研究團隊在「第三屆 BioCreative 生物文獻自動探勘」競賽中獲得第一名的佳績。**

由資訊所許鈞南研究員實驗室所研發的基因正規化系統，在一項由美國 國科會所主辦的第三屆 BioCreative 生物文獻自動探勘競賽中獲得第一名的佳績。這項競賽 的成績是在 2010 年 9 月 13-15 日於美國馬里蘭州的美國國家健康研究院 (NIH) 所 在的 Bethesda 市舉行的研討會中由主辦單位發表的。

卓越講座系列

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 Superieure	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

• 最新訊息請參閱本所網站 •

國際研究生生物資訊學程卓越講座系列

國際研究生學程卓越講座系列，第一場於 2010 年十一月十二日於中研院活動中心舉行。由資訊所特聘研究員許聞廉教授主講，講題為："Intelligence: Machine vs. Human".

(左) 副院長劉兆寧博士及主講人 (右) 許聞廉教授。

卓越講座系列

"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*

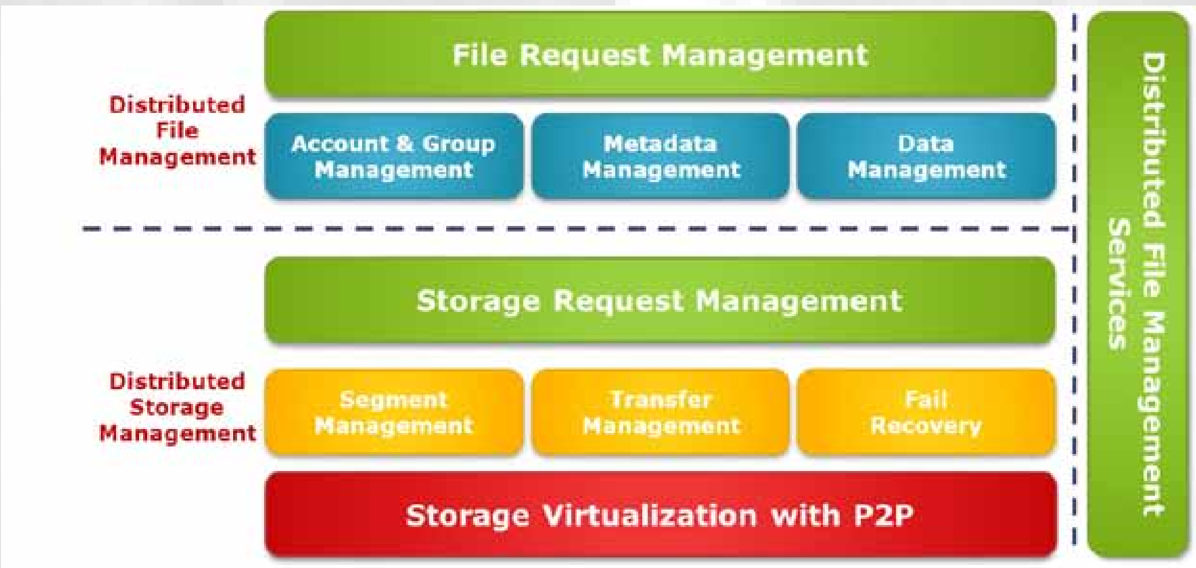
電腦系統實驗室：實現電腦的無限可能

電腦系統實驗室成立於 2009 年由游本中所長、李丕榮研究員、王建民副研究員、吳真貞副研究員、陳昇瑋助研究員及其研究群共同組成，實驗室主要研究領域跨足多核心系統、虛擬化技術及雲端運算相關系統支援及其應用。以下為其實驗室研究內容介紹：

系統設計及針對多核心系統之支援

此一團隊研究著重於提供各種嶄新的系統功能結構及系統支援，設計此類多核心系統能夠在硬體設計前先進行大量的系統層模擬。為了縮短關鍵產品上市的時間，軟體必須在模擬器上開發並同一時間硬體亦進行設計，因此加速模擬器將促進縮短設計時間及提出更好的系統設計產品。

他們利用的方法是藉著多核心系統中固有的執行緒平行度，在現有的多核心系統平台模擬新的多核心設計，為了加速此一模擬，許多現存單核心模擬加速技術已經被延伸至至多核心模擬上，其中包含了利用程式執行階段 (phase) 中相似的程式行為，將快速的功能模擬器和準確的時序精準模擬器加以整合來縮短模擬時間並取得精準的模擬結果。



提出許多特定領域之平行程式模組以幫助程式設計者。

多核心系統上的平行程式開發支援

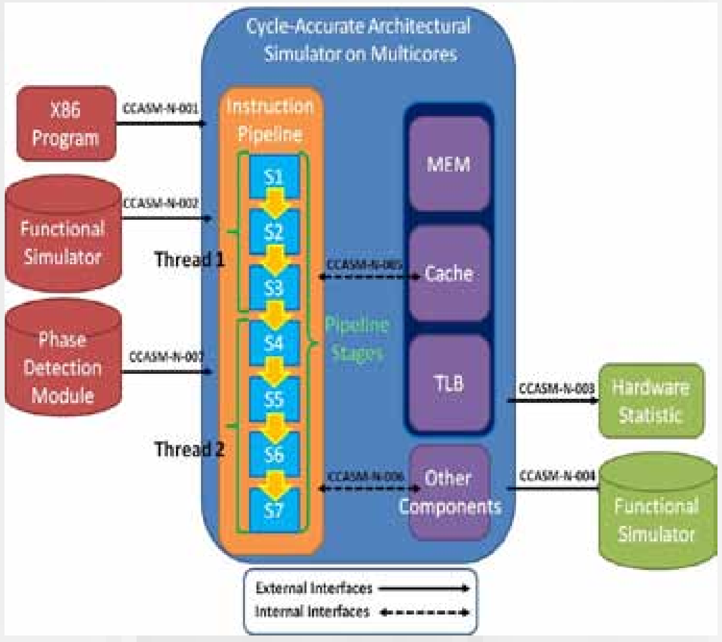
軟體應用程式是決定多核心技術能否成功的關鍵因素。然而，平行程式不僅開發艱難而且由於多執行緒在多核心系統上的互動造成難以除錯，因此能支援平行程式效能微調及在多核心系統上除錯的工具對於程式開發者能提供實際的幫助。

許多特定領域之平行程式模組被提出以幫助程式設計者將現有得應用程式移植到多核心平台上或在新的多核心平台上撰寫新的程式碼，近年來成功的平行程式模組如 Nvidia 的 CUDA 及 Google 的 MapReduce 已經在電腦繪圖及搜尋引擎中被廣泛運用，並尋求在其他領域上被運用的機會。

此團隊研究的重點包括開發新的高效率且有效支援平行應用的偵錯技術，允許程式開發者能提供對於程式的瞭解以指引編譯器能編譯出在多核心或多用途圖形處理原件上更高效能的平行程式碼。

虛擬化之系統支援

此團隊著重研究於支援系統虛擬化的核心技術，更精確而言是針對多核心系統的動態編譯技術如二進制碼轉譯及二進制最佳化。他們正尋求將其拓展從單核心至多核心系統上的方法，其中之一的主要目標是在一統一的動態編譯架構下將二進制程式碼能在多個不同的指令集架構間轉換，此一目標將在二進制程式碼階段被實現而非在高階語言虛擬機器如 Java 或 C#，因此能在現有的二進制程式碼中提供更高的透明度及可攜性。



團隊研究著重於提供各種嶄新的系統功能結構及系統支援。

支援具高度互動性應用之高效率資源運用之雲端中介軟體

由於單一實體伺服器無法容納數以百萬計的玩家同時上線，因此必須將網路虛擬環境 (NVE) 分割至數台機器上，如此一來會遇到幾點困難：第一，NVE 玩家並不是均勻分布；第二，玩家分布情況會隨時間而改變。

此團隊研究著重於提供一雲端中介軟體，使其能達到頻繁互動和高效率資源運用的目標，以提升伺服器能源使用效率和滿足即時互動需求。其中的挑戰，如虛擬環境下效能保證與公平分配、能源與資源管理、工作量預測、快速即時遷移虛擬機器、容錯處裡、記憶體內資料庫設計以及動態資料分割、複製和區域化等，皆須有創新的解決方法。

活動快訊

參訪團


在台東默默醫療服務 33 年的台東基督教醫院創院院長譚維義及其夫人譚秀麗（右上），九月二十八日於宋定懿研究員（右下）的陪同下，拜訪資訊所並會見所長。

波蘭高等教育部政務次長十一月九日來訪，與游本中所長分享波蘭高教經驗。（左至右游所長，李德財院士，Dr. Maria E. Orlowska; Dr. Leszek Grabarczyk; Mr. Yeh-Shin Chu)

Aerospace Research & Development (EOARD) in Air Force Office of Scientific Research 亞洲辦公室 Dr. David Atkinson 與 Dr. Peter Friedland 於十月二十五日拜訪資訊所，並於會中簡報。

國防部至資訊所視查國防役人員訓儲狀況，由陳克健副所長與會説明。


網路國際標準與漢字工作坊

2010 年 10 月 7 日 W3C XML Working Group 專家 Murata MaW3C XML Working Group 專家 Murata Makoto Adobe 的 Unicode 專家 Yamamoto Taro 等六位專家來台介紹網路技術與萬國碼 (Unicode) 的發展及參與經驗，並邀請國內財團法人中文數位化技術推廣基金會（中推會）、自由軟體鑄造場 (OSSF)、軟體自由協會等社群成員，與數典計畫成員進行交流。

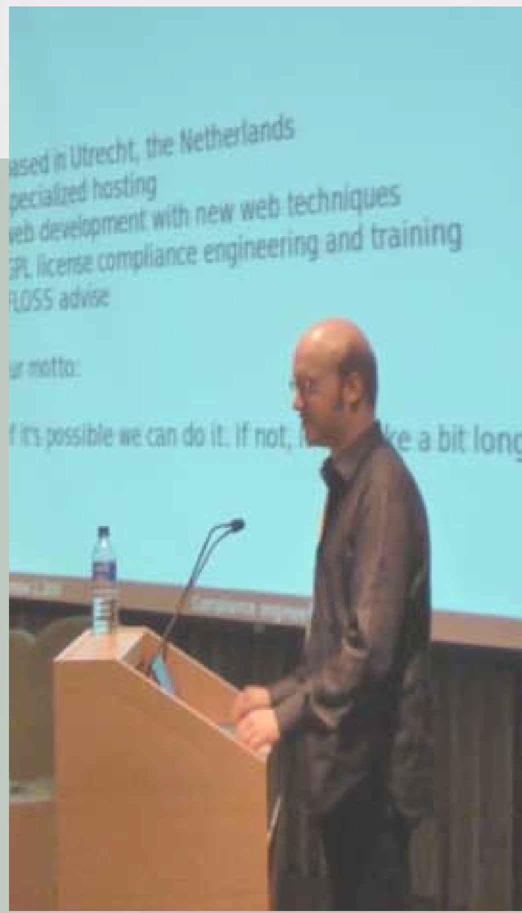
活動快訊

自由軟體工作坊

GPL 的授權規則與技術工程遵循之道 自由軟體授權應用與商業建議

自由軟體鑄造場於 2010 年十二月二日特別邀請到專門處理 GPL 侵權案件的著名組織 GPL-Violations.org 的核心成員：Harald Welte 與 Armijn Hemel 擔任本次演講的講師，探討利用 GPL 元件開發商業產品時要如何遵守授權規定。

Welte 先生本身為 GPL-Violations.org 組織的設立發起者，同時也是許多著名 GPL 元件的著作權人，近年 Welte 先生成功地在歐洲提起多件 GPL 侵權訴訟並獲得勝訴判決；而 Hemel 先生則長年在 GPL-Violations.org 組織中負責技術察核侵權利用的工作。兩位講師對於 GPL 的授權內容與相關的技術層面均具相當深入的瞭解。 詳細內容請參閱 http://www.openfoundry.org/en/workshop/details/115.



Armijn Hemel 先生闡述有關技術察核侵權之技術。

Harald Welte 先生成功地在歐洲提起多件 GPL 侵權訴訟並獲得勝訴判決。

「2010 電腦對局論壇」於 2010 年 9 月 4 日於中研院資訊所盛大舉行，促進電腦對局遊戲程式及相關領域之研究、發展、應用與交流。

2010 年院區開放活動

10 月 23 日中研院全院對外開放參觀，資訊所的各項展出吸引了眾多的民眾前來共襄盛舉，門庭若市。

各種研究成果的展現，引起熱烈的參與與回響。

資訊所門口一早就排滿了人潮。

自由軟體鑄造場
擴大軟體應用研發能量

OSSF 成員，「認真工作認真玩」是他們的座右銘。(圖中：張韻詩特聘講座教授領軍)

順應國內外自由軟體日益普及的潮流，中央研究院資訊科學研究所於民國 92 年接受經濟部工業局「自由軟體產業推動計畫」的補助，成立自由軟體鑄造場 (Open Source Software Foundry，以下簡稱 OSSF)，以期能擴大自由軟體在台灣的應用幅度與研發能量。目前經費的來源，部分來自行政院國家科學委員會的補助，部分則來自於中央研究院資訊科技創新研究中心。

OSSF 由一群對自由軟體充滿熱忱的成員所組成，並依據業務屬性，分成技術、營運、法政三組來完成每年的計畫目標。計畫主持人目前有四位，分別是李德財老師、張韻詩老師、莊庭瑞老師、以及蕭景燈老師。

OSSF 維運多元的自由軟體服務網絡，對外的服務項目包含：設置 OpenFoundry 軟體開發平台、建立軟體人才資料庫、提供軟體授權法律諮詢、匯集軟體研究資源表列、發行 OSSF 電子報，並持續舉辦相關的技術教學工作坊及校園推廣演講等。透過這些平台與服務的提供，鼓勵國內更多的程式開發者及使用者，加入自由軟體豐富多彩、樂於分享的世界。

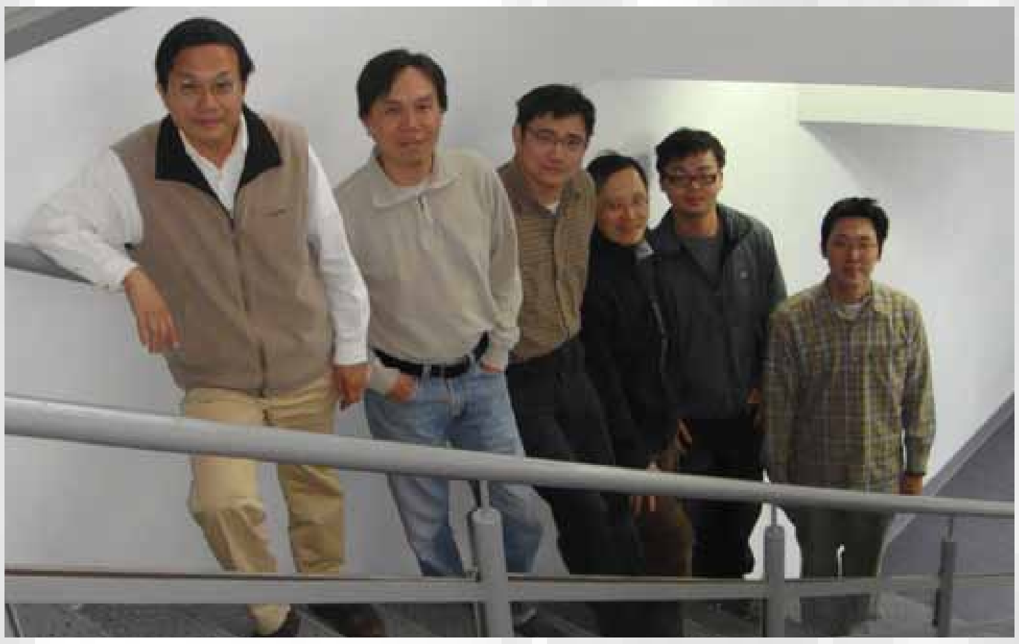
OSSF 辦公室目前座落於資訊所新館 402 室，未來亦有可能配合資創中心擴大至院內其他處所。正所謂 Work Hard Play Hard，辦公室內不定期有同仁貢獻各類好吃的家鄉名產，討論區每日不定時供應提神醒腦的自助式咖啡，還有飛鏢投擲、迴旋踢練習等運動，來幫助舒壓並激勵創意。

歡迎所內對自由軟體研究開發有興趣的同仁，能夠造訪自由軟體辦公室，認識 OSSF 活潑的工作夥伴，一同交流與分享自由軟體程式開發的經驗。

計畫網頁：http://www.openfoundry.org/
聯絡電話：02-2788-3799*1469 1477 1478
聯絡信箱：contact@openfoundry.org

自由軟體技術充電站。

疾病傳播電腦模擬
加速疾管局新流感防治決策



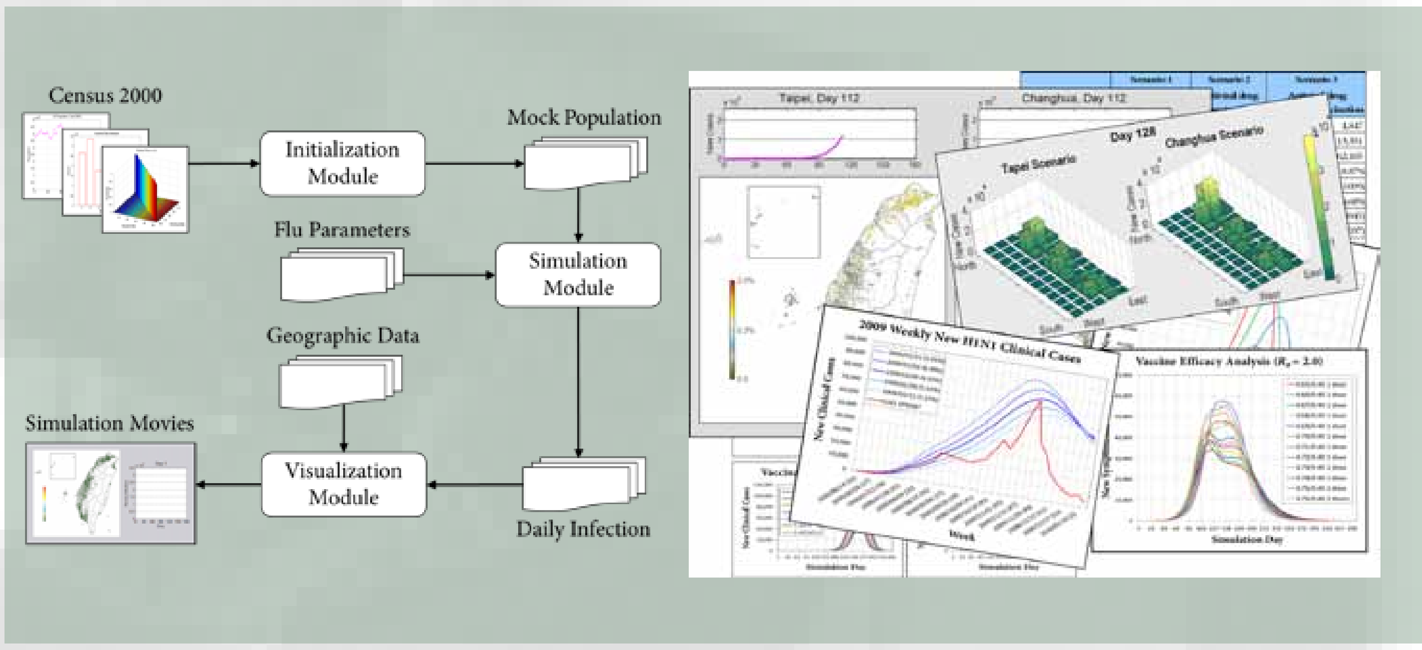
研究團隊由本所徐讚昇研究員領軍（左至右），陳淙琛先生、王大為研究員、廖純中研究員、蔡孟宗先生、張紘睿先生。

資訊所團隊大幅加快疾病傳播電腦模擬速度，提供疾管局新型流感防治政策決策參考。

世界各地先進國家在面對重大傳染性疾病散播時，均研發電腦模擬系統，以便檢視疾病在散播時的特性，提供防治參考。去年新型流感流行期間，資訊所研究團隊研發目前全球最快速之傳染病散播電腦模擬系統，提供疾管局使用，系統輸出的成果成為該局新型流感防治策略決策時重要之參考依據。本研究參與研究學者包含中研院資訊科學研究所徐讚昇博士、廖純中博士、王大為博士，他們所帶領的巨量資料運算實驗室的研究團隊，以及來自疾管局、香港大學及台灣大學的學者專家。

研究團隊表示，本研究的最大挑戰及和以前研究的不同點在於，如何把所有台灣 2 千 3 百萬人的可能日常活動及這些活動可能的空間距離關係全部放入電腦的模型中加以考量，模型的複雜度及所需的計算量都是前所未有的規模。除此之外，也利用巨量資料運算實驗室以往在個人隱私保護的研究成果，使模擬結果可以呈現所要的巨觀資訊，又不會洩漏個人資料。

考量較複雜的系統模型，模擬所需時間相對太長，無法提供及時了解目前趨勢；考量太簡單，又無法適度反映真實狀況。此系統在設計時精密到模擬系統中學生在哪一個班級，同學有沒有打疫苗都知道，可是運算速度卻比其它國家的類似系統快上約一千倍，其主要的關鍵核心在於基礎理論演算法的突破。這項研究成果也於 11 月 4 日發表在國際著名學術期刊 PLoS ONE。



由於使用這個系統，疾管局可以針對各種新型流感防治政策的各種可能狀況加以模擬，參考模擬結果，形成最經濟有效的防治決策。由於模擬速度很快，每一種狀況模擬所需時間由十幾個小時變成不到十分鐘，可以模擬比較多的假想狀況，決策的過程相對完善而有彈性。例如，如果疫苗的劑數有限，要如何挑選施打的對象即期程，使得防治效果最好。

由於是電腦模擬，有人不免擔心結果是否和真實狀況相符。以去年為例，在 2009 年 H1N1 新型流感來襲時，於 4 月 30 日即曾以此系統模擬，發現如果做好機場檢疫，可將整個流行高峰期延後 22 天，可爭取時間做好防疫準備工作。 研究團隊表示，由於此系統的運算速度快，增加模擬次數及模型精確度，因此在統計上相對的可信度也增高。

參考文獻：

http://dx.plos.org/10.1371/journal.pone.0013292

[1] M.-T. Tsai, T.-C. Chern, J.-H. Chung, C.-W. Hsueh, H.-S. Kuo, C.-J. Liau, 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. Liau and T.-s. Hsu, "Heterogeneous Subset Sampling", Proceedings of the 16th International Computing and Combinatorics Conference (COCOON), Springer-Verlag LNCS# 6196, pages 500-509, 2010.

應用資料探勘 探索發掘新世界

2008 年前往紐西蘭奧克蘭市參加第 34 屆 VLDB 國際會議做論文報告。

由 Prof. Hiroshi Motoda 頒發 PAKDD2006 Student Travel Award。

葉彌妍博士目前為中央研究院資訊科學研究所助研究員，同時合聘於中研院資訊科技創新研究中心。2009 年從台大電機研究所獲得博士學位，研究興趣主要在資料庫和資料探勘領域，有關資料串流管理與分析等重要議題。

目前除了發表兩篇論文於 IEEE TKDE (IEEE Transactions on Knowledge and Data Engineering) 期刊及其他國際會議論文，更在 2008 年於資料庫領域極具權威的會議 - VLDB (International Conference on Very Large Data Bases) 發表一篇論文。此篇論文為 2004 後迄今，唯一一篇以台灣學生身份投稿而被接受之論文。

會進入電腦科學領域從事學術研究，對葉彌妍來說不在本來人生規畫之中。高中時期，葉彌妍最喜歡也最擅長的科目是化學，著迷於各種奇妙的化學變化和分子結構，更曾經代表北一女中參加台北市化學科能力競賽獲得佳作。原本立志高中畢業後要選擇化學相關科系就讀，朝自己喜歡的領域發展；結果當年大學聯考成績意外地比預期的高，而有了「為何不給自己多一個機會探索新領域？」的想法，最後幸運地進入當年二類組的第一志願 - 台大電機系就讀，而且就一路唸到取得博士學位。

原本擔心在這個男性佔多數的科系會無法適應(當年男：女人數大約是 9：1)，但在電機系的日子卻是葉彌妍求學生涯裡最快樂的時光。因為女生相對比較少，彼此容易建立深厚的感情，並且容易得到系上師長和同學們較多的關心，反而擁有了一個良好的學習環境。在台大電機提供多元化課程之下，葉彌妍漸漸地對其中計算機組領域的課程產生濃厚興趣，也開啟了她在電腦科學領域繼續鑽研的契機。

研究所期間葉彌妍重要的人生經驗之一，就是獲得國科會千里馬計畫的補助，前往美國紐約州的 IBM T.J. Watson Research Center 訪問一年。IBM T.J. Watson Research Center 是美國知名的研發中心，匯集了來自世界各地的專家學者在此從事研究工作。在一年的訪問時光中，除了學習更多研究的技巧，葉彌妍也結識了很多來自不同國家但一樣有志於科學研究的朋友。訪問期間，受到前輩們對於研究的熱情與細心認真態度之影響，葉彌妍越來越能肯定自己從事「研究」這件事的價值：在無國界無時限的平台上，和眾多優秀聰明的人既競爭也合作，討論既存爭點、發現新問題、進一步解決問題，同時藉由推銷自己的點子，微小卻真實地影響這個世界。

當研究成果被重要國際期刊和會議接受時，埋首挑燈之辛苦頓時轉為繼續努力研究的成就感與無比動力。

目前，於中研院資訊所，葉彌妍的研究內容著重在以下三大方向：

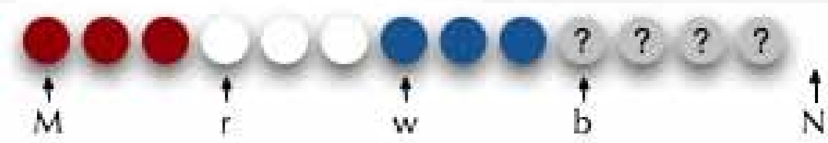
1. 資料流的管理與探勘：延續著先前博士班的研究，針對多條動態改變且快速成長的資料流，探討如何以有效率且省資源的方式，在各種條件限制下，做相似性查詢以及樣式比對等相關議題。

葉彌妍希望透過學術研究平台推銷自己的點子，微小卻真實地影響世界

2. 行動軌跡分析與探勘：現今無線通訊以及定位系統技術發達，各種物體移動的軌跡可以透過 GPS 等設備記錄下來。如何從大量且同時具備時間與空間資訊的軌跡資料中，發現有趣的物體移動模式，進而發展適時適地性服務，是極具實用性與挑戰性的研究課題。目前擬先針對市區計程車及公車等交通工具移動所記錄下來的軌跡，做各種可能的分析和探勘。

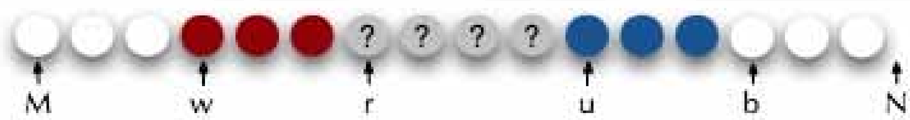
3. 社群網路分析：拜網際網路科技之賜，社群網路應用在網路世界蓬勃發展，其網路規模及複雜程度已經遠大於早期社會學家所分析的人際網絡。藉用電腦科學的力量，首先期望能找出針對異質社群網路有效收集具代表性的資料取樣方法，以供後續行為模式分析。

從荷蘭到波蘭



一個可能性是允許未知區段出現在旁邊（圖一）。

我們也可能把這未知區段放在其中兩個顏色之間（圖二）。



他們用的不變量是這樣（圖三）。

● 穆信成 ●

一個陣列裡頭每個元素都是紅、白、藍三色之一。如何藉著兩兩交換，把它們由左至右依紅、白、藍的順序排好呢？

這個「荷蘭國旗問題」由 Feijen 和 Dijkstra 提出 — 紅白藍是荷蘭國旗的顏色。（為何不是法國或南斯拉夫國旗？多半因為 Feijen 和 Dijkstra 都是荷蘭人吧！）眾所皆知，quicksort 的缺點是碰到大致上已排好的或有重複元素的陣列時，效率就變差了。若把荷蘭國旗問題用在 quicksort 中，用紅、白、藍三色分別代表小於、等於、大於 pivot 的元素，由於等於 pivot 的元素已被放在一起，應有易於處理重複元素的好處。

這種程式至少得用個迴圈，而既然有迴圈，我們得為它找個不變量。程式執行中必定有一段「未知」的區段。一個可能性是允許未知區段出現在旁邊（圖一）：

或著，我們也可能把這未知區段放在其中兩個顏色之間（圖二）：

哪個不變量比較好用呢？兩者都寫得出可運作的程式，但其中一種好寫得多。讀者要不要試試看？

然而，直到 90 年代中期之前，大家的一般認知是解荷蘭國旗問題需要太多次交換，不值得用在 quicksort 中。當時，Lee McMahon 為第七版 UNIX 寫的 quicksort 函式和其衍生版本已流傳將近二十年之久。一般做法是把大陣列先用 quicksort 分割，等切得夠小就換成 bucket sort 或 radix sort. 即使如此，Bentley 和 McIlroy 發現，對所有他們能找到的 quicksort 實作，總有個辦法做個得花長度平方時間處理的輸入。於是他們決定寫個更有效的 quicksort。

為了處理重複元素，Bentley 和 McIlroy 又回頭研究了荷蘭國旗問題。他們認為問題出在我們花了太多工夫把白元素一步步挪到中間。與其如此，不如先把白元素放到兩邊去，把陣列排成「白、紅、藍、白」。等排序完成，再把白色換回中間就可以了！他們用的不變量是這樣（圖三）：

當然，我們還需把白元素交換回中間。如果白元素多，交換次數就多。但他們指出這部份的代價可分配到演算法其他部份去，因為白元素一旦放到正確位置，便不用再被遞迴排序了。Sedgewick 和 Bentley 也證明了此種 quicksort 是熵值最佳解 — 每元素平均被比較數在熵值的常數比率之內。荷蘭國旗問題終於漸漸被用在包括 Java 與 C 的函式庫中。

只是，這麼一來，我們解的好像不是荷蘭國旗問題了。有什麼國旗是「白、紅、藍、白」四色的呢？我找不出分四條色塊的國旗。如果把顏色交換一下，倒是找到了這幅 Rypin 的旗幟。Rypin 是波蘭的一個小鎮，根據 06 年的統計，人口約一萬六千人。

參考資料

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

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

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

Shin-Cheng Mu. 荷蘭國旗問題 The Dutch National Flag Problem. http://www.iis.sinica.edu.tw/~scm/ncs/2010/10/dutch-national-flag-problem/.

The Institute of Information Science (IIS) at Academia Sinica, Taiwan, R.O.C. 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 en- couraged 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.

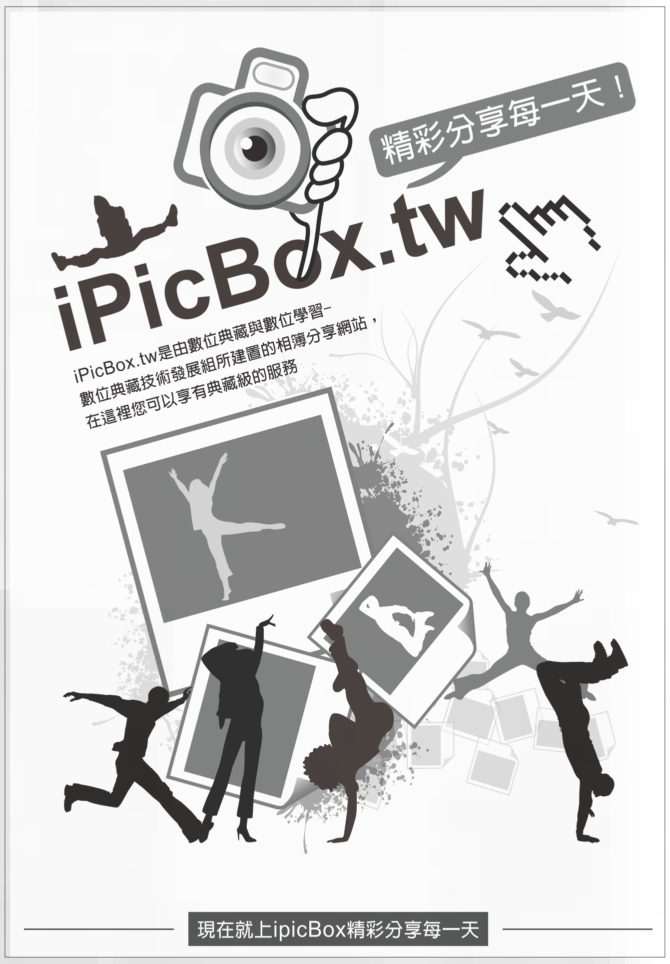
Full-time research fellows are free to set their own research directions. IIS currently has about 40 full-time research fellows and close to 300 full-time post doctoral fellows and research assistants. The areas of their current research include Systems Technology, Bioinformatics, Multimedia, Natural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

All Candidates should have a Ph.D. 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 detailed curriculum vitae, and at least three letters of recommendation to

Director
Institute of Information Science
Academia Sinica
Nankang 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 http://www.iis.sinica.edu.tw.



Institute of Information Science
Academia Sinica
128 Academia Road, Section 2, Nankang,
Taipei 115, Taiwan, R.O.C.
Tel: 02-27883799
http://www.iis.sinica.edu.tw/
Publisher: Director Pen-Chung Yew
Editors: Pei-Chi Wang and
Huey-Chyi Chris Tseng