

# 用聽的瀏覽器

## A Talking Browser

張凱揮  
Kai-Hui Chang  
國立臺灣大學電機工程  
研究所碩士班研究生  
changkh@gcn.net.tw

范潤萃  
Jun-Tsui Fan  
國立臺灣大學公共衛生  
學系大學部學生  
astrocosm@kimo.com.tw

郭斯彥  
Sy-Yen Kuo  
國立臺灣大學電機工程  
研究所教授  
sykuo@cc.ee.ntu.edu.tw

### 摘要

在本論文中提出了一個「用聽的瀏覽器」的架構並加以實作，可以將網頁上的文字資料翻譯成另一語言後轉換為自然語言輸出。對盲人、不識字者、手機使用者、企業及中國大陸皆有極為重要的應用和效益。

關鍵字：瀏覽器、自然語言處理、客語、翻譯

### Abstract

The structure of “A Talking Browser” is proposed and implemented in this paper. It can translate the text information obtained from a web page to another language, and then read it out using the target language. For blind people, illiterates, cell-phone users, companies and Mainland China, there are important applications and benefits.

Keywords: Browser, natural language processing, Hakka, translation

### 一、導論

目前網際網路已十分普遍且盛行。上網瀏覽網頁、收發電子郵件等已成為現代許多人每天必備的工作。透過各式各樣的網頁，人們可以獲取許多資訊，增加自己的視野，也擴充自己的生活及工作能力。但在一般人享用網路資訊的同時，同時也拉大了和弱勢族群間的差距。某些族群因為目前以視覺為主的瀏覽器不適用而沒有辦法上網瀏覽，因此就減少吸收新資訊的機會，使得這些族群更顯弱勢。

對目前以視覺為主的瀏覽器而言，不識字的人和盲人是沒有辦法使用的，而這兩種人正是社會上接受資訊較少的弱勢族群，需要新的科技來改善他們的生活，而不是讓他們更顯弱勢。所以需要有一個機制，讓網頁上的文字資料能夠轉為聲音唸出，使這些族群的人也能夠透過網路吸收新資訊，以拉近和一般人的差距。

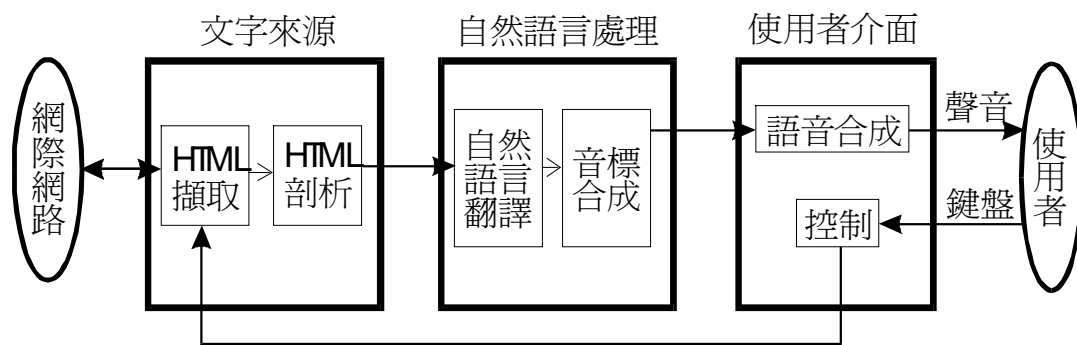
目前網路上的華文資料絕大多數是以華語撰寫而成的。對於不識字的人而

言，他們大多聽不懂華語而只會一種或兩種生活中常用的方言，此種情形在臺灣和中國大陸都很普遍[1]。因此在設計用聽的瀏覽器時尚需考慮語言之間翻譯的問題，以擴大使用的範圍。若是不需要翻譯時只需要將翻譯功能關閉以原語言讀出即可。

在本論文中，首次結合了網頁瀏覽、自然語言翻譯和讀稿等技術，合成一實用的系統，並針對華語翻譯成客語完成了初步的實作，可說是一大創新。除了用聽的瀏覽器本身的技術外，其操作的方法、使用的對象以及效益分析也是本篇論文的重點所在。

## 二、系統架構及演算法說明

系統架構如圖一所示。



圖一、系統架構

其各部份的功能詳述如下：

### 1. 文字來源

這部份可以再分為兩部份：HTML 擷取和 HTML 剖析。在 HTML 擷取的部份是要能使用 http 協定取得網頁檔案，而剖析的部份是要從 HTML 檔中取得我們所需的資訊，包括超連結和文字資料。

### 2. 自然語言處理

#### (一) 理論

這部份可以分為兩部份：自然語言翻譯和音標合成。一般而言，自然語言處理有幾個層次，包括語彙分析(lexical analysis)、文法分析(syntax analysis)以及語意分析(semantic analysis)等[2][3]。在做自然語言翻譯時，最好的方法是根據來源語言的資料建出了一句話語意的剖析樹後，再利用相反的過程使用目的語言的語意、文法和語彙庫重新產生自然語言以完成翻譯，但如此太過複雜困難。因為我們翻譯的語言和目的語言都是漢語的一支，只是分屬不同的方言，所以在語意和

文法上很少有歧異，最主要是在做語彙的轉換，所以只需要處理到語彙的層次，而利用一些文法或語意分析幫助決定一些用字即可。

在翻譯上第一步是要斷詞。中文的詞與詞之間不像英文句子一樣中間有空白，所以必需用人工智慧的方法來斷詞。所謂的斷詞就是將一句話中的每一個詞斷出來。斷詞的方法很多，在系統中主要是利用「長詞優先」為原則，從左往右尋找。在翻譯時首先以目前尚未處理的字串為目標在詞典中搜尋。如果碰到幾個詞都可以和目前字串開頭的字符合，則取其最長者代換，叫做「長詞優先」。例如句子開頭為「我們」一詞，而詞庫中有「我」和「我們」兩詞，則斷詞將從其長者斷成「我們」，而不是「我」「們」兩詞。

斷詞完後就需要詞代換。此時需要兩種語言詞互轉的資料庫，在本系統中就是華語和客語互轉的資料庫。資料庫中要儲存的是被翻譯語言在目標語言中的講法。例如華語中的「他」在客語中為「佢」這樣的資訊。若是斷出的詞在資料庫中沒有這樣的轉換，則視為目的語言中的講法和被翻譯語言相同而不翻譯，在本系統中也就是客語講法和華語講法一樣的词。例如「政府」、「臺北市」等。

詞代換時有許多情形要考慮。有時候華語中的同一個詞在客語中有不同講法，但意義相同，此時就利用機率任擇其一。有時候華語中的同一個詞在客語中有不同講法，意義也有所不同，就需要做字義分析而加以解決。這在自然語言處理上叫做字義辨異(word sense disambiguity)問題。在處理上通常是利用 Gale 提出的演算法[4]。

在文法結構上也有一些要考慮的。例如華語中「只」這個概念在客語中是在句尾加「定定」。像是「只有一個人」客語講法是「有一個人定定」。目前的做法是在一個句子中用字組比對(pattern matching)的方法找到「只」這樣的關鍵字將之刪除後在句尾加上「定定」。此時也可以用字義辨異的方法輔助以確定這裡的「只」代表的意義是可以代換成「定定」的那個意思。

在翻譯後的文字產生後就要產生音標。此時主要根據已斷的詞查含有音標的詞典，如果有則直接輸出，如果沒有該詞的資料則查單字的字典輸出音標。此時會碰到一字多音的問題，就需要根據文法分析或語意分析標定該字的詞性或字義後再加以判別。其演算法亦可以利用 Gale 的演算法，或是其他自然語言處理的演算法。如 Markov Model [5]、Trigram Model 等。

翻譯句的音標取得後就要處理連續變調的問題。所謂的變調指的是某兩種聲調的音結合成詞時音調改變的情形。在客語四縣腔中的連續變調有三種情形[6]，以實際音值表示列表於表一：

變調前	變調後
24 + 24	11 + 24
24 + 55	11 + 55
24 + 5	11 + 5

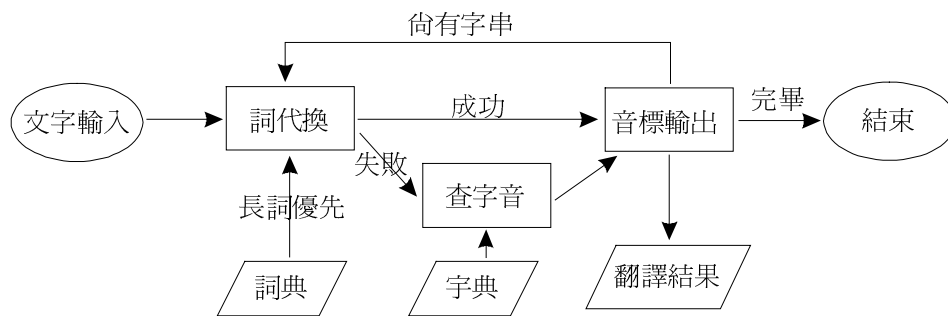
表一、客語四縣腔連續變調

例如陰天本來為 im24 thien24，經變調後即變為 im11 thien24。

變調處理上最大的問題在斷詞，因為變調時要以詞為範圍先變，之後才跨詞變調，這樣才能符合口語使用的習慣。所以前面的斷詞資訊要一直留到這裡。

## (二) 實作

因我們手邊資料庫的限制，在目前實作上和理論略有不同。主要是因為我們沒有華語詞庫，只有華語轉客語的詞庫和客語字典，所以自然語言處理的架構也有所不同。我們實作的步驟到連續變調處理前如圖二。



圖二、自然語言處理實作架構

因為沒有華語詞庫，所以斷詞是以客語詞典中有的詞為依據來斷詞。因為缺乏訓練語料庫，所以目前在一詞多義或一字多音時都是用亂數取其一，而沒有用字義辨異的演算法。至於連續變調目前為不考慮斷詞，從左而右變調。

## 3. 使用者介面

使用者介面分為語音合成和控制兩部份。在語音合成中將自然語言處理單元輸出的音標唸出，而在控制部份則是接受使用者輸入送到 HTML 擷取部份選擇要從網際網路中擷取的 HTML 檔。

因為我們在音標輸出時選擇的不是標音位的音標，而是標實際音值的音標，所以音標和語音是一對一的，因此可以直接輸出。輸出的方法可以用語音合成的技術將音標合成實際語音，而在實作上是利用台大客家社客語有聲字典中預先錄好的字音檔直接播出。該字音檔每一個檔是一個音，檔名即音標，包含變調結果後的聲音檔也有，所以只要按音標順序播出即可。

在控制部份是要接受使用者從鍵盤輸入的指令而做出對應的工作。在實作上是利用 MFC 撰寫 Windows 下的程式來達成。

## 三、實作結果

### 1. 功能說明

系統的功能最主要是將網頁中的文字部份化成自然語言唸出，主要是將網頁中的華語換成客語唸出。至於要唸的東西為何，就透過使用者介面由使用者決定。

在使用者介面上我們設計成利用鍵盤上的數字鍵加兩個功能鍵即可操控。使用者可以隨時用數字鍵選擇要進入的超連結。例如要選擇第二個超連結，輸入 2 及 enter 即可進入該超連結。此外亦可用 0 代表回上頁，00 代表到下一頁，000 代表重讀本頁，而以 ESC 結束程式。若是在手機上可利用"\*"代替 ESC，利用"#"代替 enter。

第一個進入的網頁可以利用檔案設定，其內容可隨使用者的需要而定，通常以常用網站的網址為最方便。這部份可由他人先行代為設定。

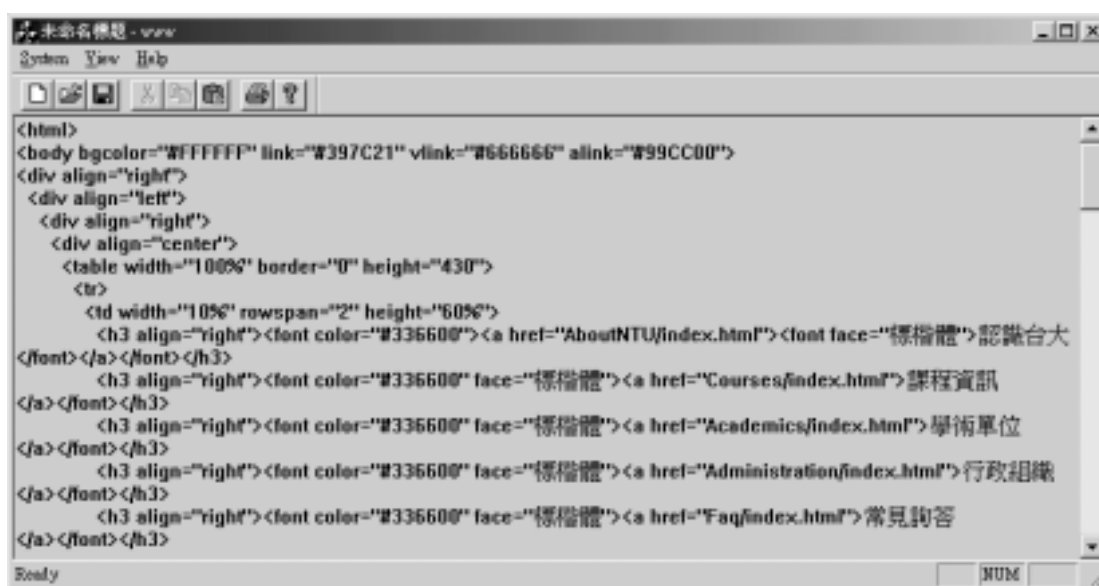
至於怎麼知道哪一個數字代表哪一個超連結的方法如下：系統在讀稿時碰到超連結會先放出一個鈴聲，之後讀出一個數字代表該超連結，接著繼續讀文字。在讀到超連結結束符號後再放出一個鈴聲表示超連結結束。

例如「請到<a href="h1.htm">台灣大學</a>的<a href="h2.htm">電機工程研究所</a>來。」這個 html 檔將讀出「請到（鈴聲）一台灣大學（鈴聲）的（鈴聲）二電機工程研究所（鈴聲）來」。之後若是按 1+enter 就會跳到 h1.htm 繼續讀出網頁。

## 2. 執行情形

因為瀏覽器的輸出結果為聲音，這裡只能利用文字描述。

程式的執行畫面如圖三。測試的網頁是台灣大學的純文字版首頁。



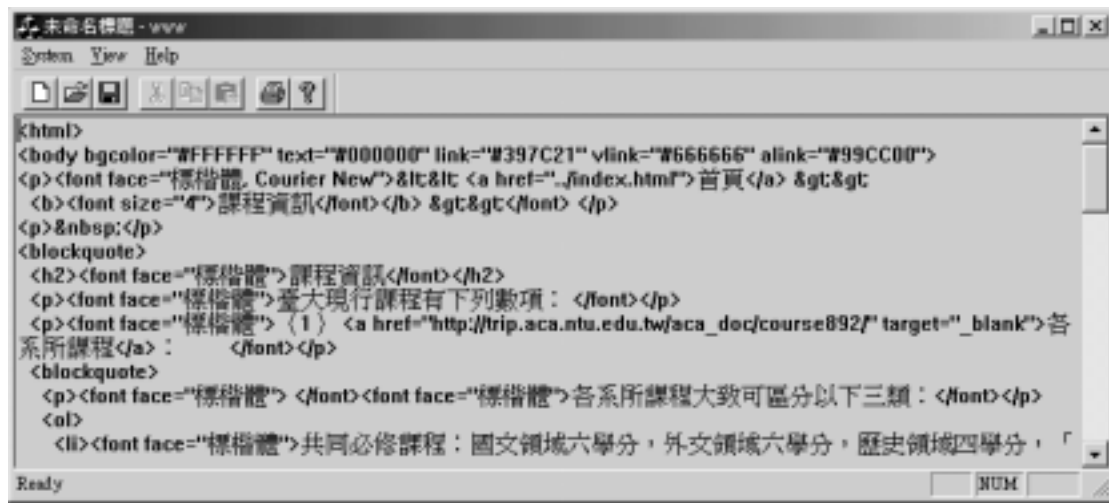
```
<html>
<body bgcolor="#FFFFFF" link="#397C21" vlink="#666666" alink="#99CC00">
<div align="right">
<div align="left">
<div align="right">
<div align="center">
<table width="100%" border="0" height="430">
<tr>
<td width="10%" rowspan="2" height="60%">
<h3 align="right"><font color="#336600"><a href="AboutNTU/index.html"><font face="標楷體">認識台大
</font></a></font></h3>
<h3 align="right"><font color="#336600" face="標楷體"><a href="Courses/index.html">課程資訊
</a></font></h3>
<h3 align="right"><font color="#336600" face="標楷體"><a href="Academics/index.html">學術單位
</a></font></h3>
<h3 align="right"><font color="#336600" face="標楷體"><a href="Administration/index.html">行政組織
</a></font></h3>
<h3 align="right"><font color="#336600" face="標楷體"><a href="Faq/index.html">常見詢答
</a></font></h3>
```

圖三、執行畫面一

畫面中顯示的文字是其 html 檔和程式除錯用的資訊，實際唸出來的聲音如下：

「(鈴聲)一認識台大(鈴聲)(鈴聲)二課程資訊(鈴聲)(鈴聲)三學術單位(鈴聲).....」

如果選擇 2+enter，進入的畫面如圖四。



圖四、執行畫面二

其唸出的內容如下：

「(鈴聲)一首頁(鈴聲)課程資訊 課程資訊 臺大現行課程有下列數項 (鈴聲)二各系所課程 (鈴聲) 各系所課程大致可區分以下三類.....」

## 四、應用及效益分析

### 1. 應用於盲人及不識字者

#### (一) 增加工作機會

目前一般人求職已經不易，對盲人及不識字者而言就更加困難了。如果利用「用聽的瀏覽器」配合語音辨識軟體，則盲人及不識字者可以用電腦寫履歷，而且透過網路寄出履歷，使求職更加方便有效率。就工作本身而言，當盲人或不識字者可以直接取得網路上各種資訊而不需要額外的幫助時，可以做的工作就更多了。

#### (二) 便利盲人和不識字者的生活

在現代社會中許多一般人視為輕而易舉的事情對於盲人或不識字者而言都是很困難的。像是買賣股票、轉帳、購買車票、查火車時刻表等。可是當盲人或不識字者有能力使用網路後，這些事將變成可能，大大便利了他們的生活。如果是上購物網站，不只前面說的事情能做到，想要買任何東西幾乎都可以完成。

#### (三) 改善盲人的教育環境

盲人學校可以透過電腦輔助教學，由學校購置電腦設備，而讓學生可以用電腦學習、上網查資料、交作業、交報告。成年盲人也可以在家裡上網學習，使其不和社會脫節。

## 2. 應用於企業

在網頁和語音輸出相關的情形下皆可使用。例如許多觀光區都有語音導覽系統。如果開發一套網頁，利用「用聽的瀏覽器」技術翻譯成各國語言，將可大大減少開發成本，也便於維護。

## 3. 應用於醫療

醫學天天都在進步，手術的方法也日新月異，病人的病症又千奇百怪，有時外科醫生突然碰到一個急需開刀的病人時，也會需要上網查詢治療的方法，根據說明一步一步執行。可是緊急時又不一定有時間看網頁搞清楚了再做。此時如果利用用聽的瀏覽器配合語音控制，就可以讓電腦把手術步驟一步一步唸出來，醫生就可以跟著一步一步做手術，挽救病人的性命。

除此之外，在偏遠落後也沒有網路的地方，若能利用電話配合用聽的瀏覽器查詢醫療網站，也往往能夠協助醫生的診斷，甚至在沒有醫生的地方自己解決一些醫療的問題，提升醫療水準。

## 4. 應用於中國大陸

中國大陸幅員廣大，文盲仍然不少，方言分歧的情形更是普遍，而沿海和內地的經濟程度差異極大。對居住於內地中不識字的人而言，很難有接受新資訊的機會，如此大大地影響著下一代的教育和生活品質。目前雖然各省的電視臺和電臺有方言節目的播放，但在山區等收視不良的地方並不一定能夠收看或收聽，而電視的價格也不是人人都負擔得起的。

如果使用「用聽的瀏覽器」，只需要一部普通功能的電腦加上一條電話線，就可以達成資訊傳播的功能。電腦可以將沿海公家機關淘汰的電腦或是回收的電腦直接拿來使用，幾乎不需成本。而電話線架設容易，在文字資料使用頻寬極少的情形下，一條電話線即可提供很多人使用「用聽的瀏覽器」。而在內容方面，可在中央成立一小組將各種資訊轉換成文字，專門讓「用聽的瀏覽器」的使用者瀏覽，或是直接瀏覽網路上的各種網頁。由於自然語言翻譯和讀稿是在使用者的電腦上完成，中央只需要輸入華語，不需要特別的人翻譯成中國所有的方言。對於平衡城鄉差距，全面提升內地生活和知識水平有極大的幫助。

目前世界各國對自然語言翻譯的研究很多，但對漢語各方言之間的研究卻很少，本論文中提出的技術正好可以用在中國內部各種不同方言和次方言間的翻

譯，十分具有價值。

## 5. 應用於手機族

目前的手機族上網常詢的資訊通常以文字為主，但在手機上的小螢幕要看這樣多的資訊是十分辛苦的，更何況有時候使用者並不能看著螢幕，例如在開車或做運動時。此時若使用「用聽的瀏覽器」就可以解決使用不便的問題。

## 五、結論

本篇論文中提出並實作了一個結合自然語言處理和網際網路技術的「用聽的瀏覽器」，而且對其使用對象和效益做出分析。對於盲人、不識字者等弱勢族群而言，「用聽的瀏覽器」更是連接資訊社會、改善他們生活的一扇窗。網際網路的應用，透過「用聽的瀏覽器」，將可更進一步發揮它的效用，深入更多人的生活。

## 感謝

感謝臺灣大學資訊工程學系自然語言實驗室陳信希教授、林川傑同學提供技術協助。

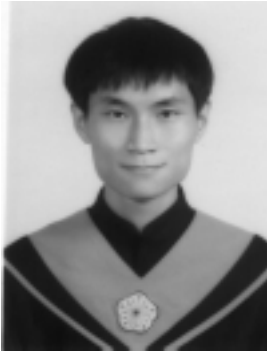
感謝臺灣大學客家研究社提供客語字典、客語詞典以及客語有聲資料庫。網址 <http://www.south.nsysu.edu.tw/group/NTUHakka>

## 參考資料

- [1] 董同龢，漢語音韻學，文史哲出版社，1963
- [2] Gerald Gazdar and Chris Mellish, Natural Language Processing in Lisp, Addison Wesley, 1989
- [3] Christopher D. Manning and Hinrich Schutze, Foundations of Statistical Natural Language Processing, The MIT Press, 1999
- [4] Gale, William A., Kenneth W. Church, and David Yarowsky, “A Method for Disambiguating Word Senses in a Large Corpus”, Computers and the Humanities, 26 pp 415-439, 1992
- [5] Jelinek, F., “Markov Source Modeling of Text Generation”, The Impact of Processing Techniques on Communications, J. K. Skwirzinski, Ed. Nijhoff, Dordrecht, 1985
- [6] 羅肇錦，台灣的客家話，臺原出版社，1990

## 作者簡歷





**Kai-Hui Chang** was born in Tainan, Taiwan on December 10, 1976. He received the B.S. degree (1999) and M.S. degree (2001) in Electrical Engineering from National Taiwan University, Taipei, Taiwan. His research interests include natural language processing, Internet technology and circuit simulation algorithms.



**Jun-Tsui Fan** was born in Taipei, Taiwan on January 4, 1977. She is majoring in Public Health for the B.S. degree from National Taiwan University. Her research interests include Chinese phonology, health care, and preventive medicine.



**Sy-Yen Kuo** received the B.S. (1979) in Electrical Engineering from National Taiwan University, the M.S. (1982) in Electrical & Computer Engineering from the University of California at Santa Barbara, and the Ph.D. (1987) in Computer Science from the University of Illinois at Urbana-Champaign. He is a Professor in the Department of Electrical Engineering, National Taiwan University since 1991. He spent his sabbatical year as a visiting researcher at AT&T Labs-Research, New Jersey from 1999 to 2000. He was the Chairman of the Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan from 1995 to 1998, a faculty member in the Department of Electrical and Computer Engineering at the University of Arizona from 1988 to 1991, and an engineer at Fairchild Semiconductor and Silvar-Lisco, both in California, from 1982 to 1984. In 1989, he also worked as a summer faculty fellow at Jet Propulsion Laboratory of California Institute of Technology. His current research interests include dependable distributed systems and networks, software reliability, and optical WDM networks.

Professor Kuo is an IEEE Fellow. He has published more than 150 papers in journals and conferences. He received the distinguished research award (1997-2001) from the National Science Council, Taiwan. He was also a recipient of the Best Paper Award in

the 1996 International Symposium on Software Reliability Engineering, the Best Paper Award in the simulation and test category at the 1986 IEEE/ACM Design Automation Conference (DAC), the National Science Foundation's Research Initiation Award in 1989, and the IEEE/ACM Design Automation Scholarship in 1990 and 1991.