

# Phát triển phần mềm chữ Nôm

Ngô Trung Việt, Viện CNTT, [vietnt@itprog.gov.vn](mailto:vietnt@itprog.gov.vn)  
Ngô Thanh Nhân, Đại học New York, [nhan@cs.nyu.edu](mailto:nhan@cs.nyu.edu)  
Trần Lưu Chương, Khu công nghệ cao Hoà Lạc, [chuongtl@itprog.gov.vn](mailto:chuongtl@itprog.gov.vn)  
Đỗ Bá Phước, Vietnamese Nôm Preservation Foundation, [jdo@usa.net](mailto:jdo@usa.net)  
Nguyễn Quang Hồng, Viện Hán Nôm, [hongnq@vol.vnn.vn](mailto:hongnq@vol.vnn.vn)  
Nguyễn Hoàng, California, [hoang.myloc@gte.net](mailto:hoang.myloc@gte.net)  
Lê Phạm Ngưng Hương, Thụy Sĩ, [lepham@dial.eunet.ch](mailto:lepham@dial.eunet.ch)  
Hà Dương Tuấn, Paris, [tuan.hd@wanadoo.fr](mailto:tuan.hd@wanadoo.fr)  
Đỗ Tuyết Khanh, Hồ Văn Tiến, Thụy Sĩ, [h\\_tien@operamail.com](mailto:h_tien@operamail.com)

## ***Tóm tắt***

*Trong gần mười năm qua, trong những cố gắng mã hoá cho chữ Việt và các chữ dân tộc chính như Chăm, Thái vào bộ mã chuẩn quốc tế Unicode, chữ Nôm đã được dành một phần quan tâm rất lớn. Việt Nam đã tham dự đều đặn các cuộc họp của Nhóm các báo cáo viên chữ biểu ý IRG, và đã đứng ra đăng cai tổ chức 2 cuộc làm việc quốc tế của nhóm này tại Việt Nam. Các nỗ lực khai thác kho chữ Nôm của dân tộc và đưa vào kho chữ chung của nhóm IRG đã được tiến hành liên tục và có nhiều kết quả. Với thông báo của Unicode vào tháng 3/2001, chính thức 9299 chữ Nôm của Việt Nam đã được đưa vào cả bộ mã Unicode 3.1 lẫn ISO 10646. Điều này mở ra những khả năng mới cho việc sử dụng chữ Nôm trên máy tính và đưa chữ Nôm trở về với việc sử dụng rộng rãi trong các gia đình.*

*Bài báo này, sau khi giới thiệu những nét chính đã đạt được trong việc thực hiện đưa chữ Nôm vào bộ mã chuẩn quốc tế, sẽ trình bày một số vấn đề và định hướng nghiên cứu trong phát triển phần mềm xử lý chữ Nôm. Những vấn đề nghiên cứu chính được đặt ra ở đây là:*

- *Phát triển chương trình bàn phím để gõ chữ Nôm vào máy tính.*
- *Xây dựng cơ sở dữ liệu phục vụ cho tự điển chữ Nôm.*
- *Xây dựng các cơ sở dữ liệu về văn bản cổ có chứa chữ Nôm dựa trên XML/Unicode.*
- *Đưa các tác phẩm chữ Nôm vào máy tính.*

- *Tổ chức Web site về chữ Nôm.*
- *Xây dựng một động cơ tìm kiếm cho Quốc ngữ và chữ Nôm.*
- *Bước đầu phát triển chương trình phiên chuyển thông minh Nôm - quốc ngữ.*

*Đây là một phần việc quan trọng cần có được sự tham gia đông đảo và rộng rãi của các chuyên viên tin học trẻ.*

## **0. Dẫn nhập**

Người Việt Nam dùng chữ Nôm làm chữ viết chính hơn 10 thế kỷ nay, cho tới năm 1920 chữ Quốc ngữ mới chính thức thay thế cho chữ Hán và chữ Nôm. Hai cuộc chiến tranh với Pháp và Mỹ đã loại trừ phần lớn những hiểu biết về thứ chữ viết này trong cộng đồng người Việt Nam. Đây là sự mất mát lớn nhất về văn hoá sau mất mát về con người và vật chất của chiến tranh. Những phát triển về kỹ thuật xử lý đa ngôn ngữ trong máy tính và mạng Internet từ giữa thập kỷ 1980's đã mở ra khả năng hệ thống hoá và trình bày (in ấn) chữ Nôm trong mạng toàn cầu, do đó đem lại khả năng phục hồi việc sử dụng chữ Nôm và giữ gìn các công trình văn hoá của người Việt Nam.

Việc phổ cập chữ Nôm đã đến lúc phải đặt thành chương trình quốc tế song song với chữ quốc ngữ. Đây là vấn đề then chốt trong việc khôi phục vốn văn hoá Việt Nam tiềm tàng trong các kho chữ Nôm ở Việt Nam (từ gia đình, làng xã, ... cho đến kho quốc gia). Hiện nay trên thế giới đã xuất hiện khả năng thành công cho việc này. Sự thành công này không thể có trong tiến độ nhanh chóng nếu trong nước tự làm lấy, mà là do sự hợp đồng khởi xướng của những người quan tâm trong nước và ngoài nước, bất kì quốc tịch nào.

Vào đầu những năm 1990, quan niệm phổ biến về việc gìn giữ chữ Nôm trong nước mới chỉ là cất giữ các văn bản cổ vào kho tư liệu quốc gia, và do đó không chống lại nổi với sự mất mát thêm do thời gian, lại càng làm hạn chế số người có khả năng tiếp cận tới chữ Nôm. Thực sự, việc gìn giữ bất kì vốn văn hoá nào chỉ có thể được thực hiện thành công nếu điều đó được gìn giữ trong nhân dân, trong việc sử dụng sống hàng ngày, chứ không phải là việc gìn giữ kiểu bảo tàng.

Nhận thức được tầm quan trọng của việc này, gìn giữ di sản văn hoá dưới dạng sống thực, một số chuyên gia người Việt trong nước và ở nước ngoài hợp tác nhau đã đặt vấn đề sử dụng kỹ thuật hiện đại giúp phục hồi sử dụng chữ Nôm. Ý tưởng đó đã được triển khai thành những nghiên cứu đề xuất đầu tiên về việc đưa chữ Nôm vào bộ mã Unicode và ISO 10646. Tiếp đó, với những nỗ lực của các chuyên gia trong nước và người Việt ở nước ngoài, tác động vào các cơ quan chủ yếu trong nước chịu trách nhiệm về vấn đề này, và dần dần có sự hợp tác thật sự tích cực. Các cơ quan có trách nhiệm trong nước đã bắt đầu dồn tài lực và nhân lực cho công tác này. Sự việc còn được đẩy lên một mức độ cao hơn khi một số công ti phần mềm quốc tế bắt đầu chú ý tới việc phát triển phần mềm cho chữ Nôm. Điều này chứng tỏ một khi có định hướng đúng và tác động đúng mức,

đúng lúc, chúng ta có thể tạo ra được phong trào phục hồi sử dụng chữ Nôm trên công cụ máy tính hiện đại.

Trường hợp giữ gìn chữ Nôm bằng việc đưa vào các bộ mã quốc tế Unicode và ISO 10646 chứng tỏ tiến trình mở cửa văn hoá có nghĩa là việc giữ gìn văn hoá VN có thể làm song song cả ngoài lẫn trong nước. Mở cửa nghĩa là đem cái của ta để ra ngoài thay vì có nghĩa "hội nhập" (dấu cái của ta, lấy cái của người). Chữ Nôm là một ví dụ mà ai cũng thấy ta cần phải đưa ra ngoài. Ai cũng biết là trình bày và phổ biến chữ Nôm hiện nay đã là điều khả thi. Tuy thế, riêng trong Việt Nam thì không có đủ nguồn nhân lực, thời gian và vật lực để làm các công việc chuẩn bị kỹ thuật cho việc phục hồi sử dụng chữ Nôm, trong khi tiến trình toàn cầu hoá xảy ra nhanh quá mức. Do đó một sự kết hợp toàn diện các tiềm năng chất xám trên thế giới mới có thể giúp cho điều này được thực hiện.

Hơn nữa, việc giữ gìn và bảo tàng các gia tài văn hoá Việt Nam phải xảy ra trên trường quốc tế. Việc sử dụng các phương tiện hành chính để bảo vệ các công trình lịch sử trong những năm qua đã cho thấy nhiều thất bại hơn thành công. Người Việt Nam phải biết sử dụng phương tiện thông tin và bảo vệ tác quyền (mà hiện nay Việt Nam vẫn chưa là thành viên của Công ước Berne) của thế giới để gìn giữ "gia tài" của Việt Nam.

Chúng ta hãy xét lại những việc đã làm được trong thời gian qua và định hướng cho những phát triển sắp tới đây có liên quan tới chữ Nôm.

## **1. Bước đầu đưa chữ Nôm vào bộ mã chuẩn quốc tế**

### **Các hoạt động xây dựng chuẩn quốc tế liên quan tới Việt Nam**

Theo tài liệu của Unicode [5], Liên đoàn Unicode được thành lập 12/1990 nhằm vào việc xây dựng và đưa vào sử dụng bộ mã chuẩn quốc tế 16 bit Unicode cho các ngôn ngữ trên thế giới. Song song với việc này, tổ chức tiêu chuẩn quốc tế ISO tiến hành xây dựng bộ mã chuẩn quốc tế 16-32 bit cho mọi loại chữ viết có trên thế giới, dưới tên gọi ISO 10646. Hai bộ mã chuẩn này đã được thống nhất với nhau và một phần quan trọng trong bộ mã này là đưa vào mã hoá cho các chữ của các nước Đông Á.

Chữ quốc ngữ của Việt Nam (với dấu thanh xem như dấu tổ hợp) đã được Liên đoàn Unicode và ISO IEC 10646 chấp nhận vào năm 1991 [1].

Cuộc họp đầu tiên của ISO về việc thống nhất chữ Hán để đưa vào bộ mã quốc tế được tổ chức tại Bắc Kinh năm 1989. Cuộc họp thứ hai của ISO được tổ chức tại Seoul tháng 2/1990 tại đó thành lập ra nhóm nghiên cứu liên hợp Trung/Nhật/Hàn Chinese/Japanese/Korean Joint Research Group (CJK-JRG). CJK-JRG đã trở thành nhóm con chính thức của ISO/IEC JTC1/SC1/WG2 và được đổi tên thành Nhóm báo cáo viên chữ biểu ý Ideographic Rapporteur Group (IRG) vào tháng 10/1993.

Việt nam bắt đầu tham dự cuộc họp CJK-JRG lần thứ 5 tổ chức tại Hawaii, đại diện của Việt Nam đã đưa ra lời mời nhóm tới họp ở Hà Nội và lời mời đã được chấp nhận.

"Trong cuộc họp thứ hai [của nhóm IRG] tại Hà Nội tháng 2/1994, IRG đã đồng ý đưa chữ Nôm của Việt Nam vào các phiên bản tương lai của Kho chữ và trật tự thống nhất Unified Repertoire and Ordering (URO) và bổ sung thêm từ điển tham khảo thứ năm cho sơ đồ sắp thứ tự này (Phụ lục E, trang 2, [5]).

Đến tháng 11/1995, "IRG đã xem xét tổng cộng 21,252 chữ biểu ý CJKV do Trung Quốc, Nhật Bản, Hàn Quốc, Đài Loan và Việt Nam đệ trình xem như sự mở rộng cho URO 2.0." [5]

### **Các hoạt động của Việt Nam**

Sự quan tâm của Việt Nam đối với việc đưa chữ Việt vào máy tính bắt đầu từ giữa những năm 1980. Vào đầu năm 1990, cơ quan chức năng có thẩm quyền của Việt Nam về việc phát triển các bộ mã cho chữ Việt là Tiểu ban chuẩn hoá mã chữ Việt do Trần Lưu Chương phụ trách. Việt Nam đã xin đăng kí làm thành viên liên hệ của tổ chức Unicode từ năm 1993. Những nghiên cứu về chữ Nôm và máy tính bắt đầu từ những nỗ lực riêng lẻ của một vài nhóm chuyên gia Việt Nam trong nước và ngoài nước.

Tại Mỹ, Ngô Thanh Nhân, Đỗ Bá Phước và Nguyễn Hoàng đã nghiên cứu đưa ra một giới thiệu về chữ Nôm của Việt Nam cho tiểu ban chữ Hán của Unicode tháng 7/1992 [2]. Theo sự uỷ nhiệm của Tiểu ban chuẩn hoá mã chữ Việt, các tác giả đã trình bày giới thiệu này cho nhóm nghiên cứu hỗn hợp Trung Quốc - Nhật Bản - Hàn Quốc (China-Japan-Korea Joint Research Group) CJK-JRG tháng 12/1992 tại Hawaii [3]. Nghiên cứu này tiếp đó được Ngô Thanh Nhân, Đỗ Bá Phước, Nguyễn Hoàng phát triển thành bộ chữ thuần Nôm (Nom Proper Code Table) NPCT 1.0 và được lấy làm bản đề nghị của Việt nam trong cuộc họp IRG#1 tại Fukuoka, Nhật Bản vào tháng 10/1993. Trong nước, nhóm nghiên cứu về chữ Nôm và tin học của Ngô Thế Long tại Viện Nghiên cứu Hán Nôm cũng bắt đầu nghiên cứu việc tin học hoá chữ Nôm.

Trong năm 1992, Tổ chức tiêu chuẩn quốc tế ISO đưa ra bỏ phiếu về bộ mã chuẩn 32 bit đa ngôn ngữ ISO 10646, phần bao hàm cả bộ mã Unicode. Hoàn cảnh lúc đó là đã có 23 phiếu ủng hộ và 8 phiếu chống, thiếu một phiếu bầu nữa là đủ tỉ lệ 3/4 để cho bộ mã này được thông qua. Đỗ Bá Phước, uỷ viên Ban kĩ thuật Unicode, đã thông báo cho Tiểu ban chuẩn hoá mã chữ Việt về vấn đề bỏ phiếu này và đề nghị Việt Nam nên bỏ phiếu ủng hộ và đồng thời yêu cầu Tổ chức tiêu chuẩn quốc tế chấp nhận đề nghị đưa chữ Nôm của Việt Nam vào bộ mã chuẩn đa ngữ này. Trưởng tiểu ban chuẩn hoá mã chữ Việt, Trần Lưu Chương, đã làm các đề nghị để cho Tổng cục tiêu chuẩn đo lường chất lượng, đại diện chính thức của Việt nam tại ISO, bỏ phiếu tán thành bộ mã này với yêu cầu chấp nhận đưa chữ Nôm của Việt Nam vào bộ mã này trong các lần phát triển tiếp sau [4].

Việc đưa chữ Nôm vào bộ mã chuẩn quốc tế bắt đầu trở thành một trong những nhiệm vụ chính mà các cơ quan có trách nhiệm của Việt Nam dần dần tham gia thực hiện. Các cơ quan phối hợp trong công tác này là Tổng cục tiêu chuẩn đo lường chất lượng, Tiểu ban chuẩn hoá mã chữ Việt (sau này chuyển sang Ban kĩ thuật Công nghệ thông tin (CNTT)), Viện Công nghệ thông tin và Viện Nghiên cứu Hán Nôm. Hai nhiệm vụ chính được đặt ra lúc này là cần phát triển bộ mã chuẩn cho chữ Nôm làm cơ sở pháp lí để yêu cầu đưa vào bộ mã chuẩn quốc tế, xây dựng phong chữ Nôm và chuẩn bị tổ chức cuộc họp của IRG - Nhóm các báo cáo viên chữ biểu ý IRG#2 tại Hà Nội tháng 2/1994. Tại cuộc họp

IRG#2 này, chữ Nôm của Việt Nam, dưới dạng tiêu chuẩn quốc gia TCVN 5773, được chấp nhận đưa vào kho chữ của IRG.

Vấn đề là trong số 2357 chữ Nôm được đề nghị trong NPCT và TCVN 5773 thì khoảng 500 chữ Nôm thống nhất được với các chữ đã có trong vùng CJK (Trung Quốc, Nhật Bản, Hàn Quốc) của bộ mã ISO 10646 theo qui tắc thống nhất chữ biểu ý, nhưng lúc đó chưa có cột dành riêng cho Việt Nam trong bộ mã này. IRG mặc dầu chấp nhận các đề nghị chữ của Việt Nam, đã không có thẩm quyền trong việc cấp cột riêng cho Việt Nam trong cả hai bộ mã Unicode và ISO 10646. Ban kỹ thuật CNTT đã phải cử các đoàn đại biểu Việt Nam tham dự các phiên họp làm việc của WG2 trong năm 1994 để chính thức yêu cầu cột V cho Việt Nam và cho đến năm 1995, yêu cầu này được chấp thuận: Việt Nam có một cột V riêng cho các chữ Nôm trùng hình chữ Hán, được thống nhất với các chữ đã được Trung Quốc, Nhật Bản, Hàn Quốc đề nghị, trong vùng CJK của bộ mã này. Các vấn đề khác nảy sinh là Việt Nam cần có một tiêu chuẩn khác về chữ Nôm để đưa vào thêm các chữ Nôm thống nhất được với các chữ biểu ý đã có và đồng thời nghiên cứu vét nốt các chữ thuần Nôm để xin đưa vào phần bộ mã mở rộng của Unicode và ISO 10646. Công việc này được tiến hành suốt từ năm 1995 cho tới nay thông qua việc Việt Nam tham gia tích cực vào các hoạt động của Nhóm báo cáo viên chữ biểu ý IRG và đưa các chữ Nôm của Việt Nam vào kho chữ của IRG.

## 2. Tham gia các hoạt động của nhóm IRG

Mục tiêu trực tiếp nhất mà nhóm IRG muốn đạt tới là cùng nhau xác lập một kho chữ biểu ý được tiêu chuẩn hoá và được thống nhất công nhận từ phía các thành viên có chủ quyền trong nhóm IRG đối với loại chữ này. Dựa trên kết quả làm việc của nhóm IRG, Tổ chức ISO quốc tế sẽ cấp mã cho từng chữ một.

Để thực hiện mục tiêu này, *hoạt động* của nhóm IRG bao gồm một loạt các công việc sau đây mà mỗi thành viên trong nhóm đều phải góp phần thực hiện:

### 2.1. Phân chia *lớp hạng ưu tiên* để đưa chữ biểu ý vào kho.

Trên đại thể, nhóm IRG đã phân biệt các lớp hạng sau đây:

- a. Ưu tiên đưa vào kho và cấp mã đối với những chữ hiện đang còn được sử dụng phổ biến ở các nước trong khu vực đồng văn Hán tự nhưng chưa có trong CJK. Đó là những chữ thuộc lớp A, được mã hoá vào vùng CJK Extension A trong mặt phẳng đa ngữ cơ bản (Basic Multilingual Plan - BMP).
- b. Thứ đến là những chữ đang được các nước trong khu vực sử dụng nhưng không thường xuyên và ít phổ biến. Đây là những chữ thuộc lớp B, sẽ được chuẩn hoá và đưa vào kho CJK Extension B. Những chữ Nôm do người Việt tự tạo trên là thuộc vào lớp hạng B này.
- c. Những chữ cổ hoặc thuần túy địa phương, chỉ đôi khi xuất hiện trong những văn bản hạn hẹp, được xếp vào lớp hạng C. Những chữ thuộc lớp hạng này sẽ được

xem xét đến để nhập kho CJK Extension C, sau khi đã lập mã xong cho các lớp chữ A và B.

## 2.2. Xác định nguồn chữ để đưa vào kho chữ chung.

Các thành viên của IRG đã nhất trí như sau:

- a. Đối với chữ Hán nguyên gốc, lấy *Tự điển Khang Hi* (in lần thứ 7 tại Bắc Kinh) làm nguồn chính. Và tự điển này cũng là căn cứ để tham chiếu các tự điển và tư liệu làm nguồn khác của tất cả các nước thành viên IRG.
- b. Đối với chữ biểu ý được sử dụng ở các quốc gia, thì lấy các chữ có trong các chuẩn quốc gia hay chuẩn công nghiệp của từng nước. Sau đó yêu cầu này được mở rộng bằng việc chấp thuận những chữ biểu ý có mặt trong trong các bộ tự điển sau đây, đã được IRG thừa nhận:
  - *Hanyu Dazidian* của Trung Quốc (in lần thứ nhất).
  - *Daikanwa Jiten* của Nhật Bản (in lần thứ 9).
  - *Daejajeon* của Hàn Quốc (in lần thứ nhất).
  - *Tự điển chữ Nôm* (Sài Gòn, 1971) và *Bảng tra chữ Nôm* (Hà Nội, 1976) của Việt Nam.
- c. Đối với những chữ biểu ý được các quốc gia tự tạo ra để dùng cho tiếng bản ngữ, thì ngoài những tự điển ở (b), còn có thể sử dụng thêm các nguồn tư liệu khác nhau. Việc bổ sung này được đặt ra từ đầu năm 1998, khi kho chữ được mở rộng sang lớp hạng B và C. Danh sách các tư liệu này do từng nước đề nghị và được nhóm IRG thống nhất chấp nhận vào tháng 5-1998, gồm 107 tên tư liệu, trong đó có 40 tác phẩm bằng chữ Nôm do Việt Nam đề nghị.

## 2.3. Xác định đối tượng chữ được đưa vào kho chữ chung

Cần phải phân biệt mấy khía cạnh khác nhau về đối tượng được nhận vào kho:

- a. *Chữ và biến thể* của chữ. Các hình chữ biểu ý đại diện cho các ngữ tố khác nhau là đối tượng hàng đầu của việc nhập kho chữ chung. Không có sự phân biệt chữ phồn thể với chữ giản thể, mặc dầu các chữ cùng nghĩa nhưng hình chữ khác nhau là có thể được xem như các chữ khác nhau. Căn cứ để coi các chữ khác nhau là dựa vào qui tắc thống nhất (unification) hình chữ của chúng. Nếu các chữ do các nước đề nghị trông hơi khác nhau nhưng vẫn tuân thủ qui tắc thống nhất này về cấu tạo chữ, thì được coi là như nhau và được cấp cho một mã duy nhất.
- b. *Bộ thủ* của chữ. Một danh sách 214 bộ thủ truyền thống (theo *Khang Hi tự điển*) được coi như thuộc kho chữ và có mã riêng. Bên cạnh đó còn một danh sách gồm 116 các biến thể hay dị thể của các bộ thủ (CJK Radical Supplement) cũng được coi thuộc kho và được cấp mã riêng.

## 2.4. Xác lập mã nguồn và thể thức lập bảng chữ biểu ý chung.

- a. *Mã nguồn*. Mỗi nước thành viên IRG, sau khi thu thập các chữ biểu ý cần lập bảng chữ đề nghị nhập kho, trong đó mỗi chữ phải được mã hoá theo nước mình (gọi là mã nguồn - source code) để làm cứ liệu tham chiếu. Mặt khác mỗi chữ phải được gắn với bộ thủ cùng số nét của nó để lấy đó xác định ra vị trí của chữ trong toàn bộ kho. Ví dụ: với chữ Nôm **ba** • chúng ta cấp mã (V-code) là V0-354F để tham chiếu vào các thông tin nội bộ khác về chữ này (cách đọc, xuất xứ...). Chữ này được xác định là thuộc bộ "nhất" (vì chữ "tam" thuộc bộ "nhất"), với 6 nét (không tính bộ thủ), như vậy chiếu vào *Khang Hi tự điển* thì chữ Nôm này sẽ được định vị là 0078.131 (nghĩa là nó được xếp tiếp theo sau chữ thứ 13 ở trang 78 của tự điển này).
- b. *Thế thức lập bảng*. Tất cả những chữ do các nước thành viên đề nghị, sau khi đã được IRG chấp thuận, thì được sắp xếp vào một bảng chữ chung, và tiến hành quá trình thống nhất hoá (unification) các chữ, theo qui định riêng về việc thống nhất. Thực chất, đó là việc tạo ra một cuốn tự điển của toàn vùng, thu nạp toàn bộ các chữ biểu ý đã qua chuẩn hoá và sắp xếp thống nhất. Tập chữ này được gọi là SuperCJK, được bắt đầu soạn thảo từ đầu năm 1998, qua mười ba lần điều chỉnh và bổ sung, đến năm 2001 tới bảng *SuperCJK 13*, tổng số chữ biểu ý được thu nạp và cấp mã quốc tế là 70.205 chữ (trong đó có 9.299 chữ do Việt Nam cung cấp). Trong tập SuperCJK này, các chữ được xếp thứ tự theo bộ thủ và theo mã số Khang Hi đã được định vị, và mỗi chữ đều được ghi rõ mã nguồn, mã quốc tế.

### 3. Các hoạt động liên quan tới chữ Nôm

#### Xây dựng các đề nghị đưa chữ Nôm vào kho IRG

Trong suốt quá trình tham gia chương trình hoạt động cùng nhóm IRG, các chuyên gia Việt Nam đã lần lượt soạn thảo 4 bảng chữ Nôm chuẩn để cung cấp cho IRG với ý thức tranh thủ đưa chữ Nôm vào kho chữ chung ISO/IEC-10646 càng nhiều càng tốt. Các bảng chữ Nôm chuẩn của Việt Nam lần lượt được đánh số là V0, V1, V2, V3 (và hiện đang gấp rút hoàn thành thêm bảng V4) [6]:

**V0:** Đây là bộ mã chuẩn 16-bit chữ Nôm do Ngô Thanh Nhân, [James] Đỗ Bá Phước và Nguyễn Hoàng soạn thảo, Ban kỹ thuật CNTT đề nghị, Tổng cục Tiêu chuẩn Đo lường Chất lượng xét duyệt và được Bộ Khoa học Công nghệ và Môi trường công bố năm 1993 (TCVN 5773:1993). Với bộ mã chuẩn chữ Nôm này, Việt Nam đã đăng kí đưa chữ Nôm vào kho chữ của IRG trong cuộc họp thường kì lần thứ 2 tại Hà Nội đầu năm 1994. Bảng chữ Nôm chuẩn này gồm 2357 chữ, trong đó có hơn 70% là chữ Nôm không trùng hình với chữ Hán.

**V1:** Bộ mã chuẩn 16-bit chữ Nôm do Viện Hán Nôm biên soạn, Ban Kỹ thuật Công nghệ thông tin đề nghị, Tổng cục Tiêu chuẩn Đo lường Chất lượng xét duyệt và Bộ Khoa học Công nghệ Môi trường công bố năm 1995 (TCVN 6056:1995). Bộ chữ này thu thập tất cả 3361 chữ Nôm hoàn toàn trùng hình với chữ Hán (có mặt trong *Khang Hi tự điển*). Nhóm nghiên cứu tin học thuộc Viện Hán Nôm chịu trách nhiệm chính trong việc soạn thảo bộ chữ này và vẽ phông chữ 24x24 bitmap.

Các chuyên gia Ken Lunde, Nguyễn Quang Hồng, Ngô Trung Việt, [James] Đỗ Bá Phước, Ngô Thanh nhân đã rà xét đối chiếu lại giữa V0 và V1, thấy có 38 chữ trùng lặp phải loại bỏ. Số còn lại của cả hai bảng chữ, có 3897 chữ được chính thức đưa vào vùng CJK thuộc BMP của ISO/IEC-10646 từ giữa năm 1997. Sau đó còn có thêm 128 chữ Nôm cũng được cấp mã trong bảng chữ mở rộng CJK Extension A của IRG. Như vậy là từ cuối năm 1997, có 4025 chữ Nôm (trùng hình với chữ Hán) được đề nghị cấp mã ISO/IEC-10646. Kết quả này được khẳng định tại cuộc họp thường kì IRG#10 tại thành phố Hồ Chí Minh tháng 12/1997. Đến đây, nhóm IRG cũng kết thúc giai đoạn I của chương trình, chủ yếu dành cho việc xem xét cấp mã cho các chữ biểu ý thuộc lớp hạng A.

Từ năm 1998, nhóm IRG chuyển sang giai đoạn II là giai đoạn xem xét và cấp mã cho các chữ thuộc lớp hạng B và C. Đây là dịp chữ Nôm thuần Việt được chấp nhận hàng loạt vào kho chữ chung. Không gian mã hoá mới dành cho giai đoạn II là mặt phẳng số 2 song song với BMP rất lớn (trên 64 000 vị trí mã), cho nên hầu như không phải tranh chấp ưu tiên cấp mã nữa. Nguồn chữ không nhất thiết phải là các tự điển, mà có thể là nhiều ấn phẩm khác nhau. Theo đó, các chuyên gia ở Viện Hán Nôm và Viện Công nghệ Thông tin đã khẩn trương soạn thảo thêm 2 tập mã chữ Nôm chuẩn để làm việc với nhóm IRG trong giai đoạn mới:

**V2:** Tập chuẩn chữ Nôm này do Viện Nghiên cứu Hán Nôm xét duyệt và gửi tới nhóm IRG ngày 17/4/1998. Cũng như hai tập chữ V0 và V1, nguồn chữ mà V2 thu thập vẫn là hai bộ tự điển chữ Nôm đã nói ở trên (Sài Gòn, 1971 và Hà Nội, 1976). Tập chữ Nôm này gồm 3371 chữ, trong đó có 844 chữ là bổ sung vào những vị trí tương ứng trong Extension A (trùng với chữ Hán và chữ các nước khác, đã được cấp mã), còn lại 2527 chữ (hầu hết là chữ thuần Nôm) được đưa vào Extension B. Phong chữ 96x96 bitmap do Viện Nghiên cứu Hán Nôm xây dựng.

**V3:** Tập chuẩn này do Nguyễn Quang Hồng soạn thảo, Ngô Trung Việt tạo mã nguồn, Viện Nghiên cứu Hán Nôm xét duyệt và gửi tới nhóm IRG ngày 7/5/1998. Tất cả 849 chữ hầu hết là thuần Nôm được thu thập vào tập chuẩn này là lấy từ 40 tác phẩm và tự điển khác nhau. Phong chữ 96x96 bitmap do Viện Nghiên cứu Hán Nôm xây dựng.

Sau 2 năm làm việc tiếp theo (đầu năm 1998 đến cuối năm 2000), có thêm 5274 chữ Nôm nữa (chủ yếu từ V2, V3 và phần còn lại của V0, V1) được thu nạp vào Extension A và Extension B của IRG để cấp mã trong ISO/IEC-10646. Tổng cộng lại, như trên đã nói, đến cuối năm 2000 trong bảng tổng hợp SuperCJK 13 có mặt tất cả **9299** chữ mang tên Việt Nam (với kí hiệu V0, V1, V2, V3). Trong đó có gần 50% là những chữ Nôm không trùng hình với chữ Hán. Vào tháng 3/2001, 9299 chữ Nôm được chính thức công nhận vào Unicode 3.1 <<http://www.unicode.org/unicode/reports/tr27/>> <<http://www.unicode.org/charts/>> và ISO/IEC-10646.

Tuy nhiên, nhóm IRG không muốn dừng tại đây, mà tiếp tục khuyến khích các nước thành viên cung cấp thêm các tập chữ thuộc lớp hạng C, và cả những chữ còn bỏ sót, để

lập thêm Extension C. Thành viên IRG, Nguyễn Quang Hồng và Ngô Trung Việt, hiện đang khẩn trương hoàn thành tập chuẩn V4, gồm 1049 chữ thuần Nôm, trong đó có hơn 70 chữ Nôm Tày (không trùng với chữ Hán và chữ Nôm Việt, thu thập từ công trình nghiên cứu văn hoá Tày của GS Nguyễn Văn Huyền). Font chữ 128x128 bitmap do Viện Nghiên cứu Hán Nôm thực hiện. Như vậy, nếu mọi việc tiến triển như dự kiến, thì cuối năm 2002, tổng số chữ Nôm của Việt Nam hội nhập vào kho chữ biểu ý biểu ý của khu vực và quốc tế với mã của ISO/IEC-10646 sẽ lên tới con số hơn 10 000 chữ (khoảng 50% là chữ thuần Nôm, không trùng hình với chữ biểu ý của các nước khác).

Cũng cần nói thêm rằng, mặc dù hầu như toàn bộ số chữ Nôm đã và đang đi vào kho chữ biểu ý quốc tế là được thu thập từ các nguồn tư liệu chữ Nôm tiếng Việt, song có khá nhiều chữ Nôm Tày là trùng hình với chữ Nôm Việt (và cả với chữ Hán). Bởi vậy, nếu đối với những chữ trùng hình như thế cũng tiến hành xác lập các âm đọc tương ứng trong tiếng Tày, thì cùng với hơn 70 chữ thuần Nôm Tày trong V4 đang được đăng kí cấp mã, ở một mức độ đáng kể cũng có thể lợi dụng vốn chữ Nôm Việt đã có cho tiếng Tày.

### **Xây dựng font cho chữ Nôm**

Với việc tham dự các hoạt động của IRG, các đại diện của Việt Nam đã có tiếp xúc làm việc và khuyến khích một số tổ chức nước ngoài phát triển font chữ và phần mềm cho chữ Nôm.

Tại Nhật Bản, Viện Mojikyo đã đầu tư công sức làm font cho các chữ biểu ý trong toàn vùng, nói riêng là làm font cho chữ Nôm. Viện đã tổ chức một buổi lễ trao tặng bộ font này cho Viện Hán Nôm vào tháng 10/2000 tại Tokyo Nhật Bản.

Tại Trung Quốc, công ti Dynalab nổi tiếng về làm font và phần mềm cho chữ biểu ý cũng đã bắt đầu phát triển các bộ font cho chữ Nôm.

### **Thành lập tổ chức Nền tảng Gìn giữ chữ Nôm**

Trên thế giới, nhiều tổ chức và cá nhân đã quan tâm tới việc giúp đỡ khôi phục và phát triển chữ Nôm.

Tại Mỹ, Tổ chức Nền tảng gìn giữ chữ Nôm - Vietnamese Nôm Preservation Foundation <[www.nomfoundation.org](http://www.nomfoundation.org)>, một tổ chức không vụ lợi, đã được thành lập năm 1999 để góp phần thúc đẩy sự phục hồi của chữ Nôm. Tổ chức này nhằm mục tiêu thiết lập ra một quỹ để góp phần gìn giữ các văn bản Nôm và cung cấp việc truy nhập vào chữ Nôm cho tất cả các độc giả tiếng Việt hiện đại. Do đó mở ra cánh cửa vào quá khứ văn hoá của Việt Nam. Mục tiêu cụ thể là làm cho mọi người có thể đọc được văn bản Nôm qua máy tính, xuất bản các tài liệu và nghiên cứu về chữ Nôm, và cung cấp các dịch vụ thư viện dựa trên máy tính cho các thư viện trên thế giới hiện đang có vấn đề trong việc xử lí tin học hoá kho tài liệu chữ Nôm: thư viện quốc gia Paris, thư viện Vatican và các tổ chức chính ở Mỹ, Trung Quốc, Nhật, Anh và Hà lan.

### **Xuất bản sách, tài liệu có chữ Nôm**

Năm 2000 tại Mỹ, cuốn thơ Hồ Xuân Hương dịch sang tiếng Anh -- *Spring Essence: The Poetry of Hồ Xuân Hương* -- của nhà thơ John Balaban đã được xuất bản và gây tiếng

vang lớn, đặc biệt trong cộng đồng người gốc Việt. Lần đầu tiên thơ của Hồ Xuân Hương được in cả bằng chữ Nôm, chữ quốc ngữ và chữ tiếng Anh qua chế bản điện tử trên máy tính. Điều này mở ra khả năng nhiều tác phẩm chữ Nôm nữa cũng sẽ được xuất bản bằng chữ Nôm trên máy tính.

Riêng trong nước, nhiều từ điển và tài liệu về chữ Nôm vẫn đều đặn được xuất bản, tuy nhiên vẫn chưa được thực hiện trên máy tính.

## 4. Các mục tiêu khôi phục việc dùng chữ Nôm

Để thực hiện được việc đưa chữ Nôm vào máy tính phục vụ cho sử dụng đại trà, chúng ta có nhiều mục tiêu cần đạt tới dần từng bước [8].

1. Mục tiêu thứ nhất là tiến hành *mã hoá chữ Nôm* vào trong bộ mã quốc tế chung, Unicode và ISO 10646.
2. Mục tiêu thứ hai là *xác định rõ các qui cách vào và ra thông tin chữ Nôm* để định hướng cho việc triển khai các sản phẩm phần mềm xử lý chữ Nôm.
3. Mục tiêu thứ ba là phát triển các phần mềm giúp cho việc *xây dựng cơ sở dữ liệu lưu giữ tất cả các văn bản, tác phẩm hiện có trong kho lưu trữ về chữ Nôm*. Đồng thời với điều này là việc *tổ chức và duy trì các trang Web giới thiệu về các kho lưu trữ này, tổ chức các thư viện điện tử về chữ Nôm*. Một khía cạnh quan trọng là ưu tiên vận động cho việc đưa chữ Nôm vào các font để phổ biến rộng rãi trên máy tính.
4. Mục tiêu thứ tư, bên cạnh việc tiến hành các mục tiêu trên là phải *xây dựng và hình thành một đội ngũ chuyên viên mới về chữ Nôm và tin học* để duy trì chữ Nôm dạng điện tử và đưa vào máy tính tất cả các tư liệu, tài liệu có liên quan tới chữ Nôm.
5. Mục tiêu thứ năm (gắn liền với giáo dục, đào tạo và phổ cập máy tính) là dần *hình thành một xu hướng nhiều người dân có thể truy nhập và sử dụng kho tư liệu chữ Nôm trên máy tính* khi tham khảo các tư liệu cổ. Qua đó, chữ Nôm trở thành gần gũi với cuộc sống hàng ngày của mọi người.

Như vậy những vấn đề nghiên cứu chính về kĩ thuật CNTT được đặt ra ở đây là:

1. Phát triển các bộ phông cho chữ Nôm
2. Phát triển chương trình bàn phím để gõ chữ Nôm vào máy tính.
3. Xây dựng cơ sở dữ liệu chữ Nôm.
4. Xây dựng các cơ sở dữ liệu về văn bản cổ có chứa chữ Nôm dựa trên XML/Unicode.
5. Đưa các tác phẩm chữ Nôm vào máy tính.

6. Tổ chức Web site về chữ Nôm.
7. Xây dựng một động cơ tìm kiếm cho Quốc ngữ và chữ Nôm.
8. Bước đầu phát triển chương trình phiên chuyên thông minh Nôm - quốc ngữ.

## 5. Phát triển phần mềm cho chữ Nôm

### 5.1 Phong chữ Nôm

Với số lượng hơn 10 000 chữ Nôm được chuẩn hoá và cấp mã quốc tế, đã có font chữ dạng bitmap như vậy, hoàn toàn có thể xúc tiến việc chế tạo ra các phong chữ Nôm dạng TrueType để ứng dụng cho các phần mềm máy tính.

Trong khi các tổ chức, các công ti tin học và công nghệ thông tin ở nước ta chưa thực sự lưu tâm đến những vấn đề như thế, thì từ nhiều năm nay, một số tổ chức và công ti nước ngoài đã quan tâm đến những triển vọng như vậy đối với chữ Nôm, song song với quá trình hội nhập chữ Nôm vào kho chữ biểu ý biểu ý của khu vực và quốc tế. Đến cuối năm 2000 hai tổ chức nước ngoài tự nguyện thực hiện vẽ phong cho chữ Nôm: Hội Văn Tự Kính ở Tokyo, Nhật Bản (Viện Mojikyo) và Công ti DynaLab của Đài Loan (đóng tại Thượng Hải). Toàn bộ 9299 chữ Nôm có mặt trong SuperCJK đã được cả hai cơ sở này vẽ phong đầy đủ theo các chuẩn mực đặt ra cho chữ biểu ý nói chung. Với điều kiện kỹ thuật và kinh nghiệm vẽ chữ còn rất hạn chế ở nước ta hiện nay, khó lòng có thể thực hiện được tốt một công việc đồ sộ và phức tạp như vậy. Ngoài việc bộ font cho Viện Mojikyo tặng chúng ta có thể tùy ý sử dụng nhưng chưa thuận tiện trong việc dùng thì vấn đề còn lại là thương lượng có được quyền sử dụng bộ font do Dynalab làm.

### 5.2 Bàn phím cho chữ Nôm

Sau khi đã có được những bộ phong chữ Nôm hoàn chỉnh rồi, thì công việc phức tạp đặt ra cho các chuyên gia chữ Nôm và chuyên gia tin học là làm sao có chương trình phần mềm để gõ vào chữ Nôm cũng như lưu trữ và tìm kiếm chữ Nôm như hiện nay người ta đã sử dụng được đối với chữ Hán. Một hướng xử lý thích hợp đối với chữ Nôm trong công việc này là sử dụng âm đọc của chữ Nôm theo chữ Quốc ngữ hiện hành để làm "chìa khoá" mở gọi một chữ nào đó khi cần thiết. Tính phức tạp của cách xử lý này là mỗi chữ Nôm thường có nhiều âm đọc khác nhau, và ngược lại, một âm đọc có thể tương ứng với nhiều chữ Nôm khác nhau. Xác lập cho đầy đủ và chính xác mối tương ứng giữa âm và hình của từng chữ Nôm là việc không đơn giản chút nào, mặc dù không phải là không thực hiện được (tương tự như người ta đã làm việc này đối với chữ Hán).

Một giải pháp bàn phím có thể có cho chữ Nôm là gõ vào âm theo quốc ngữ rồi tra bảng đối chiếu để lôi ra tất cả các chữ Nôm ứng với âm đọc này. Tiếp đó cần chọn bối cảnh nhỏ nhất của việc sử dụng âm đọc đó để xác định ra chữ Nôm. Ví dụ, từ âm đọc là đá, có các bối cảnh : nước đá, cục đá, cái đá, cho ... một đá, viên đá, đá lông nheo, đánh đá, bóng đá, đá cầu,... Bước cuối cùng là người sử dụng chọn đúng chữ Nôm theo ngữ cảnh đó để đưa vào văn bản.

Dĩ nhiên cũng có thể tìm kiếm những cách xử lí khác nữa để lập "chìa khoá" tra tìm chữ Nôm, như căn cứ vào bộ thủ chẳng hạn, mặc dù cách này có thể còn rối ren hơn nhiều.

### **5.3 Cơ sở dữ liệu về chữ Nôm**

Do tính phức tạp của mối quan hệ nhiều-nhiều giữa chữ quốc ngữ và chữ Nôm, cần thiết phải thiết lập ra một cơ sở dữ liệu về chữ Nôm. Cơ sở dữ liệu này gần như là nền tảng cho một cuốn tự điển chữ Nôm, và theo một nghĩa nào đó, cũng rất gần với cuốn từ điển chữ Việt cổ. Đây là chỗ ngã ba đường để thiết lập cầu nối giữa các văn bản cổ trong chữ Nôm và chữ quốc ngữ hiện đại, cầu nối giữa vốn văn hoá cổ và khả năng công nghệ hiện đại; cầu nối giữa các chữ biểu ý trong vùng và chữ la tinh hiện đại. Trên thực tế, việc tổ chức in ấn cuốn Tự điển chữ Nôm của nhóm nghiên cứu Viện Hán Nôm do Gs Nguyễn Quang Hồng chủ biên sẽ là bước đầu tiên hướng tới việc xây dựng một cơ sở dữ liệu chữ Nôm như vậy.

Cơ sở dữ liệu về chữ Nôm này có thể được mở rộng để trở thành một bảng đối chiếu giữa các chữ Hán có trong Unicode (và ISO 10646) và chữ quốc ngữ. Hiện tại trong Unicode đã có hơn 25 000 chữ Hán. Khi có bảng đối chiếu Unicode - quốc ngữ thì chúng ta cũng có thể xây dựng một từ điển đa ngữ Hán-Anh-Nhật-Hàn-Việt, lấy các chữ trong Unicode làm cơ sở. Đây sẽ là một tài liệu tốt cho việc giao lưu văn hoá dựa trên Web và Internet với các nước đồng văn cùng có truyền thống dùng chữ biểu ý.

Cơ sở dữ liệu về chữ Nôm mặt khác có thể tiếp nối với cơ sở dữ liệu chữ Việt để hình thành một kho tàng tra cứu phong phú về chữ Việt cả trong lịch sử lẫn hiện đại. Cũng cần để ý là các chữ Nôm diễn tả cho cách quan niệm và cách đọc của người Việt. Do đó có thể coi như các chữ Nôm đã được mã hoá là đã mang phần lớn việc mã hoá cho các cách nói của tiếng Việt. Việc nghiên cứu về phát âm tiếng Việt như vậy thực chất đã có được cách mã hoá cho khá nhiều từ.

### **5.4 Cơ sở dữ liệu về văn bản cổ chữ Nôm dựa trên XML/Unicode**

Các văn bản cổ viết trong chữ Nôm hiện đang được lưu giữ ở nhiều nơi trên thế giới. Một phần lớn các văn bản này đang được lưu trữ tập trung tại Viện Nghiên cứu Hán Nôm, số khác có rải rác trong các gia đình, miếu mạo, đình chùa, bia đá... trên cả nước. Một số lớn các văn bản Nôm khác nằm trong thư viện Toà thánh Vatican, các thư viện quốc gia của Pháp, Anh, Mĩ, Nhật Bản, Trung Quốc... Cần thiết phải có một hệ thống phần mềm cho phép tổ chức ghi lại nội dung của tất cả các văn bản cổ hiện đang có được vào một hệ thống cơ sở dữ liệu quốc tế để phục vụ cho việc nghiên cứu của mọi người trên thế giới. Với những tiến bộ kĩ thuật hiện nay, có thể tổ chức công việc này trong XML/Unicode.

Trong mỗi gia đình, các bản gia phả cũ còn để lại đa phần được viết bằng chữ Hán và chữ Nôm. Do đó có một nhu cầu lớn trong việc phổ cập dùng chữ Nôm cho các gia đình để cất giữ gia phả trên máy tính và để trao đổi, chấp nối các dòng họ...

Mặt khác việc thiết lập quan hệ với các thư viện trên thế giới để có thể lấy về các văn bản cổ trong chữ Nôm cũng là điều cần thiết.

### **5.5 Web site về chữ Nôm**

Tất cả những tài liệu đã thu thập được dưới dạng điện tử về các văn bản chữ Nôm cần được bố trí tại một Web site để có thể cho phép nhiều người truy nhập tới trong quá trình sử dụng và khai thác. Web site này có thể tổ chức lưu giữ

- cơ sở dữ liệu về các chữ Nôm
- thư viện điện tử về các tài liệu có chữ Nôm,
- những hướng dẫn việc học chữ Nôm.
- các kĩ thuật và phần mềm liên quan tới chữ Nôm

Web site nên được tổ chức cả trong lẫn ngoài nước.

### **5.6 Động cơ tìm kiếm cho Quốc ngữ và chữ Nôm**

Để thuận tiện cho tra cứu tìm kiếm thông tin về chữ Nôm, cần phải xây dựng một động cơ tìm kiếm cho cả quốc ngữ lẫn chữ Nôm dựa trên Unicode.

### **5.7 Chương trình phiên chuyển thông minh Nôm - quốc ngữ.**

Việc phiên chuyển từ những văn bản Nôm sang quốc ngữ là một vấn đề lớn và từ trước tới nay mới chỉ do các học giả biết chữ Nôm thực hiện. Một thực tế là chữ Nôm được đặt ra để diễn tả cách nói của người Việt, cho nên cùng một chữ viết, tùy ngữ cảnh sẽ có thể đọc thành nhiều âm khác nhau, không có một tương ứng cố định dứt khoát. Điều đó gây khó khăn cho những người mới học chữ Nôm, mặc dầu có thể biết chữ quốc ngữ. Một chương trình phiên chuyển thông minh có thể giúp đỡ rất nhiều trong việc tham khảo các cách đọc và ghi khác nhau qua các thời đại, được thể hiện trong các văn bản cổ.

## **6. Kết luận**

Các vấn đề nghiên cứu về chữ Nôm thực sự mở ra nhiều chiều hướng phát triển dùng kĩ thuật phần mềm hiện đại trên thế giới. Một mặt nó giúp cho chúng ta tiếp cận và sử dụng các kĩ thuật tiên tiến nhất, chuyển nhanh sang việc triển khai dùng bộ mã Unicode. Mặt khác cũng giúp cho việc phục hồi chữ Nôm, vốn là công cụ chuyên tải nền văn hiến xưa kia của cha ông chúng ta. Công việc này sẽ không thể nào làm được nếu thiếu sự cộng tác chặt chẽ của các chuyên gia cả trong lẫn ngoài nước cũng như sự tham gia của các thế hệ trẻ.

## **Tài liệu tham khảo**

1. A proposal for standard Vietnamese character encodings in a unified text processing framework," by [James] Đỗ Bá Phước, Ngô Thanh Nhân and Nguyễn Hoàng. *Computer Standards & Interfaces* 14 (1/1992):3-10.

2. *Introduction to Nom - a Vietnamese script - for computer character encoding*, Ngô Thanh Nhân, [James] Đỗ Bá Phước and Nguyễn Hoàng. Paper presented at the Unicode Han Subcommittee Meeting, May 7, 1992 at Xerox Corporation, Palo Alto, CA. Document No. HCS 92-3-2 and 3.
3. *Nom Proper Code Table, version 1.1*, Ngô Thanh Nhân, [James] Đỗ Bá Phước and Nguyễn Hoàng. Presented at the Fifth CJK-JRG Meeting, Nov 30-Dec 2, 1992. Honolulu, HI. Document No. CJK 5-10-CJK 5-14.
4. Tiêu chuẩn Việt Nam: *Công nghệ thông tin, các văn bản và tài liệu kỹ thuật*, Ban kỹ thuật công nghệ thông tin, TCVN/JTC1, Hà Nội 1994.
5. *The Unicode Standard, Version 2.0*, The Unicode Consortium, Addison-Wesley Developer Press, 1997.
6. Các tập mã chữ Nôm của Việt Nam cung cấp cho Nhóm IRG từ 1994 đến cuối 2000: NPCT 2.1, TCVN 5773:1993, TCVN 6056:1995, VHN 1: 1998, VHN 2: 1998
7. Ngô Thanh Nhân, *A review of dictionary indexing and lookup methods for ideographic scripts*, Hanoi, 14-17/7/1998 (<http://www.cs.nyu.edu/~nhan/vsic98.pdf>).
8. Nguyễn Quang Hồng, Ngô Thanh Nhân, Đỗ Bá Phước, Ngô Trung Việt. Chữ Nôm: văn hoá cổ truyền và thông tin hiện đại. *Tạp chí Ngôn ngữ*, Hà Nội, 1999, N4.