

VỀ BÀN PHÍM MÁY TÍNH CHO CHỮ VIỆT

Ngô Trung Việt, James Do
Ngô Thanh Nhân & Hoàng Công Thành

Viện Công nghệ thông tin

I. Các nguyên tắc xây dựng bàn phím chữ Việt

Bàn phím chữ Việt cần phải đáp ứng một số đòi hỏi có tính chất nguyên tắc phù hợp với quan điểm chung về phát triển công nghệ thông tin ở Việt Nam. Các quan điểm chính của việc phát triển công nghệ thông tin ở Việt Nam là:

- Quan điểm hệ thống mở, dựa trên cơ sở chuẩn được lựa chọn phù hợp với xu thế tiến hoá của công nghệ thông tin trên thế giới. Quan điểm này đưa tới việc cấu tạo bàn phím chữ Việt cần tuân thủ theo các chuẩn quốc tế hiện có về thông tin.

- Phát triển công nghệ thông tin ở Việt Nam trong giai đoạn hiện nay nhằm vào các ứng dụng thông tin ở Việt Nam trong mọi hoạt động kinh tế, xã hội và văn hóa. Điều này yêu cầu bàn phím chữ Việt phải phản ánh được đặc thù của chữ Việt và bảo đảm khả năng biểu diễn đầy đủ cho chữ Việt đồng thời phải đủ đơn giản để mọi người bình thường dễ dàng dùng được chữ Việt trên máy tính.

Các quan điểm có tính chất nguyên tắc trên quy lại thành các yêu cầu cơ bản đối với việc thiết kế bộ bàn phím chuẩn chữ Việt là:

- Tính đầy đủ: tức là cách bố trí và cách tổ hợp các phím có trên bàn phím phải biểu diễn được cho mọi hình chữ Việt.

- Tính nhất quán: tuân thủ các yêu cầu của các chuẩn quốc tế khi cấu tạo bàn phím và cách mã hóa ký tự.

- Tính tương hợp: dùng được chữ Việt cùng với chữ Anh trên các loại bàn phím phổ biến. Đặc biệt, yêu cầu này dẫn tới việc cần có cờ báo hiệu bàn phím đang ở trạng thái tiếng Việt hay tiếng Anh.

- Tính dễ sử dụng: để cho đại đa số người Việt có thể dùng được công nghệ thông tin, bàn phím chữ Việt cần phải có kiến trúc thỏa đáng, gần với thói quen viết của người Việt.

II. Cài đặt bàn phím chữ Việt: các mức logic và vật lí

Bàn phím là công cụ con người dùng để đưa thông tin vào máy tính. Bàn phím cho ngôn ngữ nào cũng phải phản ánh được đặc thù của ngôn ngữ đó. Nói riêng bàn phím phải cho phép đưa vào máy tính mọi con chữ có trong bảng chữ cái của ngôn ngữ. Từ yêu cầu về tính tương hợp rõ ràng rằng bàn phím chữ Việt cần dựa trên kiến trúc có sẵn của các bàn phím phổ biến trên thế giới. Vấn đề thiết kế riêng một bộ bàn phím cho chữ Việt chưa được xem xét tới ở đây do không đáp ứng được yêu cầu tương hợp.

Theo chuẩn ISO và ECMA-115 về các bàn phím quốc tế, mọi bàn phím đều được phân chia thành các mức độ cài đặt logic và vật lí. Do vậy việc xây dựng bàn phím chữ Việt cũng cần tuân thủ nguyên tắc này. Nhiệm vụ cơ bản của mọi bộ bàn phím là chuyển tín hiệu do người sử dụng tác động trên bàn phím thành các mã hiệu ứng với một đơn vị cơ sở trong bảng mã của ngôn ngữ. Vậy mức độ cài đặt logic cho bàn phím xác định quy tắc đặt tương ứng giữa các đơn vị cấu thành bảng chữ cái của ngôn ngữ với các phím logic của bàn phím. Mức độ cài đặt vật lí cho bàn phím xác định cách bố trí cụ thể các phím trên bàn phím.

Bàn phím QWERTY là một cách thể hiện cho bộ mã ASCII 7 bit, trong đó mọi kí tự hình chữ trong bộ mã này đều được đặt tương ứng với một phím bấm hoàn toàn xác định. Bàn phím AZERTY là một cách cài đặt cho bộ mã ASCII mở rộng để tả được cho mọi chữ Pháp. Để xây dựng bàn phím cho chữ Việt với tiên đề xử dụng các loại bàn phím QWERTY hay AZERTY có sẵn, cũng cần xác định ra cách cài đặt logic và vật lí cho bàn phím chữ Việt trên nền những bàn phím này.

Cách cài đặt logic cho bàn phím chữ Việt sẽ xác định ra cách phân tích các đơn vị chính tả chữ Việt thành các thành phần biểu diễn được theo bàn phím mã hoá theo ASCII 7 bit. Cách cài đặt logic này cũng còn bao gồm việc tạo lại mã đúng cho kí tự chữ Việt từ các tín hiệu thu được trên bàn phím do người dùng đưa vào. Mặt khác, cách cài đặt logic còn bao gồm cả việc tái tạo lại hình dáng đích thực của chữ Việt có mang dấu thanh thể hiện trên màn hình.

III. Đặc thù cấu tạo cách viết từ tiếng Việt

Theo TCVN 5712 thì bảng chữ cái tiếng Việt bao gồm 33 chữ cái, thêm 7 chữ cái mới nếu so với số các chữ cái trong bảng chữ tiếng Anh:

a ă â b c d đ e ê f g h i j k l m n
o ô ơ p q r s t u ư v w x y z

(các chữ cái mới thêm vào là: ă, â, đ, ê, ô, ơ, ư). Ngoài ra, tiếng Việt còn có thêm 6 thanh với 5 thanh có mang cách kí hiệu riêng:

dấu huyền, dấu hỏi, dấu ngã, dấu sắc, dấu nặng.

Như vậy so với bảng chữ cái tiếng Anh thì bảng chữ cái tiếng Việt có nhiều hơn 12 thành tố. Điều này cũng có nghĩa là không thể tồn tại tương ứng một giữa hai bảng chữ Việt - Anh. Nói cách khác, chỉ riêng các bàn phím hiện có cho tiếng Anh không đủ để diễn tả cho các chữ cái trong bảng chữ tiếng Việt, nếu không đưa ra một cách thức nào đó biểu diễn cho một chữ Việt bằng tổ hợp của vài phím trên bàn phím cho tiếng Anh.

Theo cách viết tay quen của đại đa số của người Việt Nam thì trong một từ, tất cả các chữ latin cơ bản được viết ra trước hết. Sau đó các dấu trên nguyên âm và dấu thanh mới được bổ xung thêm để tạo nên đúng chữ Việt. Như vậy trong cách viết chữ Việt đã có sự phân tích các chữ cái Việt thành tổ hợp của các chữ cái latin và các dấu nguyên âm, dấu thanh. Con chữ đúng trong bảng chữ Việt không được tạo ra ngay khi viết mà phải đợi đến giai đoạn cuối cùng, bổ xung thêm các dấu phụ thì mới thành hình. Các dấu phụ này bao gồm dấu trên nguyên âm của các nguyên âm thuần Việt và dấu thanh. Theo quan điểm cách viết như vậy có ba loại dấu nguyên âm: dấu mũ của â, ê, ô, dấu mũ ngược của ă và dấu râu của ơ, ư. Đặc thù này dẫn tới một số cách phân tích chữ Việt thành các thành phần nhỏ hơn, có thể biểu diễn được theo bộ mã ASCII. Tuy nhiên theo quan điểm này ta thiếu mất sự nhất quán với cấu tạo bảng chữ cái tiếng Việt, trong đó đơn vị nhỏ nhất của bảng chữ là các chữ cái, không có sự phân tích nhỏ hơn.

Giải pháp chính thống theo cấu tạo của bảng chữ Việt dẫn tới việc phải bổ xung thêm 7 phím logic cho 7 nguyên âm thuần Việt và 5 phím logic cho 5 dấu thanh. Do yêu cầu phải sử dụng được bàn phím chữ Việt trên các bàn phím phổ biến hiện tại, QWERTY và AZERTY, nên 12 phím logic trên phải được bố trí thêm vào cấu trúc hiện tại của bàn phím có sẵn. Giải pháp không chính thống dựa theo cách viết thông thường chữ Việt đưa tới cách bố trí các phím có sẵn của bàn phím phổ biến kiêm luôn chức năng của các dấu nguyên âm và dấu thanh: cần tối thiểu 7 phím mang ngữ nghĩa đặc biệt. Cả hai giải pháp này đều dẫn tới việc tiến hành các xử lí chặn bắt tín hiệu gõ vào từ bàn phím để biến đổi thành mã cho chữ Việt và xử lí hiện thị chữ lên màn hình. Nói cách khác, đều phải tiến hành mã hóa và giải mã cho việc tạo kí tự Việt theo các mã của bàn phím, mã ASCII.

Trong trường hợp đã có cách biểu diễn chữ Việt bằng một số tổ hợp phím thông thường, dù chính thống hay không, thì vấn đề nữa cần phải xem xét tới là cần có

cờ báo hiệu trạng thái bàn phím đang ở trong chế độ gõ tiếng Việt hay tiếng nước ngoài. Thực chất vấn đề là ở chỗ chúng ta đã làm quá tải bàn phím QWERTY hay AZERTY để có thể biểu diễn cho chữ Việt. Do đó, để phân biệt cách thông dịch tín hiệu được gõ trên bàn phím, cần phải có dấu hiệu nào đó của hệ thống xác định chế độ làm việc hiện tại của bàn phím đang trong ngôn ngữ nào. Đây là một tín hiệu thông báo cho cả hệ thống lẫn người dùng cho nên nó phải vừa được quy định ở mức mã hiệu hệ thống lẫn mức cài đặt nhận biết được cho người dùng.

4. Mã hóa cách tạo chữ Việt từ bàn phím quốc tế (bàn phím logic)

4.1 Giải pháp chính thống

Giải pháp chính thống dựa trên bảng chữ cái tiếng Việt: mã hóa các đơn vị chính tả Việt theo các phím trên bàn phím quy ước. Xuất phát từ bảng chữ tiếng Việt và so với bảng chữ tiếng Anh, các đơn vị chính tả Việt có thêm là:

ă, â, đ, ê, ô, ơ, ư,
dấu huyền, dấu hỏi, dấu ngã, dấu sắc, dấu nặng

Trong TCVN 5712, các đơn vị chính tả này đã được gán cho các mã là

Ă	A1	ă	A8	dấu huyền	b0
Â	A2	â	A9	dấu hỏi	B1
Ê	A3	ê	A0	dấu ngã	B2
Ô	A4	ô	AA	dấu sắc	B3
Ơ	A5	ơ	AB	dấu nặng	B4
Ư	A6	ư	AC		
Đ	A7	đ	AD		

Vấn đề còn lại là thiết lập tương ứng giữa các đơn vị chính tả này với một phím logic của bàn phím. Điểm xuất phát của ta là từ bàn phím quốc tế phổ biến, mọi vị trí có sẵn đều đã được gán cho một ký tự của bộ mã ASCII 7 bit theo các mức cài đặt bậc 1 và 2. Nay cũng bàn phím đó còn phải thể hiện thêm các chữ tiếng Việt, cho nên tương ứng giữa các đơn vị chính tả Việt với các bàn phím logic chỉ có thể thực hiện được thông qua việc làm quá tải phím bấm. Điều này có nghĩa là dùng các phím cũ nhưng bổ xung thêm các chức năng mới. Các chức năng mới có thể được cài đặt ở các mức từ 1 tới 4. Do yêu cầu đơn giản trong xử dụng bàn phím chữ Việt, các phím được làm quá tải này cần giữ vai trò phím sử dụng chính cho đơn vị chính tả Việt, tức là các chức năng trước đây của chúng cần được chuyển sang mức cài đặt 3 hay 4 cho bàn phím, nhường mức cài đặt 1, 2 cho chữ Việt.

4.2 Giải pháp không chính thống

Giải pháp không chính thống dựa trên cách phân tích các đơn vị chính tả Việt thành các thành phần nhỏ hơn, mã hóa được theo các phím của bàn phím ASCII.

Khi đó sẽ không có các phím tương ứng với các đơn vị chính tả Việt, tức là người sử dụng bắt buộc phải nhớ cách mã hóa chữ Việt theo kiểu tạo chữ để gõ vào máy. Điều đó dĩ nhiên là một nhược điểm khó sử dụng của giải pháp kỹ thuật này đối với đa số người sử dụng thông thường. Tuy nhiên một khi đã dùng quen thì nó lại tỏ ra rất tiện lợi và nhanh chóng.

4.2.1 Kiểu bộ mã ASCII 7 bit

Cách mã hóa dựa trên bảng mã ASCII 7 bit thực hiện mã hóa các chữ Việt theo chữ cái latin, dấu nguyên âm và dấu thanh được mã hóa độc lập nhau, dựa trên các kí hiệu có sẵn hay gần giống về hình dạng trong bảng mã ASCII. Khi đó các nguyên âm thuần Việt sẽ được mã hóa theo hai phím: phím nguyên âm latin và phím dấu nguyên âm đi kèm. Dấu thanh trong vẫn được mã hóa riêng và được đặt kèm nguyên âm dự định cho mang dấu. Sự khác đặt giữa các cài đặt cụ thể cho cách mã hóa này chỉ ở mức độ bàn phím vật lí, chọn kí tự nào trong bộ mã ASCII biểu diễn cho các dấu nguyên âm và dấu thanh. Nhược điểm chính của cách mã hóa này là tốn thời gian và số lần gõ phím để tạo ra một chữ thuần Việt. Hơn nữa các phím dấu thường nằm vào các vị trí không thuận lợi cho việc sử dụng khá thường xuyên. Ưu điểm của nó là có thể tạo ra dạng nhìn dễ nhận biết cho chữ Việt biểu diễn theo ASCII 7 bit, do đó có thể dùng được cho các mạng truyền thông 7 bit.

4.2.2 Kiểu bộ mã TELEX 5 bit

Cách mã hóa dựa trên bảng mã TELEX 5 bit bao gồm việc các nguyên âm và dấu thanh đều được mã hóa theo các chữ cái latin theo quy định thống nhất:

dấu huyền - chữ f,	dấu mũ - 2 lần kí tự nguyên âm,
dấu hỏi - chữr,	dấu mũ ngược - chữ w
dấu ngã - chữ x,	dấu râu - chữ w
dấu saes c - chũs,	
- dấu nặng - chữj	

Đặc điểm của cách mã hóa này là giúp có được tốc độ gõ vào, bàn phím rất nhanh vì việc bố trí các phím dấu đã được chọn vào các vị trí khá thuận lợi cho ngón tay gõ. Nhược điểm chính của nó là cần phải có thời gian để người sử dụng quen với cách mã hóa các con chữ Việt.

Một biến thể của cách mã hóa này người dùng gõ dấu thanh vào phần sau từ bên cạnh việc gõ dấu thanh ngay sau nguyên âm mang dấu. Hơn thế nữa, người dùng có thể thay đổi các thanh trên từ mà không phải sửa nguyên âm mang dấu thanh kể cả thanh không (tức là không có dấu thanh). Khả năng logic trong việc thay đổi thanh trên từ này có thể được cài đặt cho mọi kiểu mã hóa logic chữ Việt đã trình bày ở trên theo các bàn phím quốc tế.

4.2.3 Kiểu chữ máy Olimpia

Cách mã hóa theo kiểu máy chữ Olimpia bao gồm việc mọi chữ Việt đều được mã hóa thành chữ latin cơ sở kết hợp với dấu nguyên âm gắn kèm dấu thanh. Một số phím được chọn ra để mã hóa cho tổ hợp các dấu: dấu mũ và dấu huyền, dấu mũ và dấu sắc.. Cách mã hóa này chỉ thuận tiện cho người đánh máy đã quen với loại máy chữ này. Ngoài ra, việc bố trí các tổ hợp dấu nguyên âm và dấu thanh là không thỏa đáng và mang nhiều dấu ấn của các ràng buộc kĩ thuật từ thời các máy chữ cơ khí.

4.2.4 Kiểu con chữ dựng sẵn

Cách mã hóa theo con chữ dựng sẵn: dùng các chức năng cài đặt ở mức 3, 4 để tạo ra chữ Việt có mang dấu thanh. Có nghĩa là dùng các phím "Alt" hay "option" đi kèm theo các phím khác để tạo chữ Việt có mang dấu. Nhược điểm lớn nhất của cách mã hóa này là tổ hợp các chữ Việt có mang dấu thanh quá lớn nên gây khó khăn cho người sử dụng trong việc ghi nhớ các phím bấm.

V. Mã hóa theo bàn phím vật lí

Tất cả các phương án mã hóa theo bàn phím logic trên đây đều đáp ứng được yêu cầu biểu diễn đầy đủ cho chữ Việt. Yêu cầu tương hợp cho các loại bàn phím quốc tế cũng được thỏa mãn do việc xuất phát điểm của ta là từ các bàn phím quốc tế, bên cạnh việc bảo toàn mọi tính chất của bàn phím quốc tế, chúng ta chỉ tìm cách làm quá tải thêm nó để có thể diễn tả được cho các chữ Việt. Yêu cầu về tính nhất quán cũng được thỏa mãn vì ngoài phần đã được bảo toàn của bàn phím quốc tế, đã tuân thủ các chuẩn quốc tế về bàn phím, phần chúng ta phát triển thêm cũng dựa trên các chuẩn quốc tế.

Sau khi xác định được quy tắc mã hóa các chữ Việt theo các phím bấm trên bàn phím quốc tế và các quy tắc phân tích, tái tạo chữ Việt từ các tín hiệu bàn phím, vấn đề nữa cần phải đề cập tới là xác định cách bố trí vật lí cho các phím đó trên bàn phím. Việc này cần được tiến hành với mục tiêu đáp ứng cho yêu cầu dễ sử dụng đối với đại đa số người dùng vốn chưa quen thuộc nhiều với tin học. Việc này mặt khác đòi hỏi phải thông qua các số liệu thống kê về tần suất của các con chữ Việt và thanh để xác định các tổ nguyên âm và thanh nào hay xảy ra nhất từ đó được bố trí vào các vị trí thuận tiện nhất cho ngón tay đó.

Các số liệu thống kê về tần suất của các con chữ Việt được tóm tắt như sau. Trong số 6979 vần hiện có trong tiếng Việt (xem Nhan, 1984), tần suất xuất hiện của các nguyên âm là:

ă	577
â	583
ê	410

yê / iê / ya / ia /	417 (ya, ia = 60)
ô	518
ơ	583
uô / ua / uê	208 (ua = 63)
ư	248
ưạ / ươ	303 (ưạ = 74)

3,847 (hoặc 3,724 nếu bỏ bớt ia, ya, ua)

55.12% 53.36%

tức là hơn một nửa các văn hiện có đều mang dấu thanh. Có thể kết luận rằng nếu ta chấp thuận một chuẩn cho bàn phím có dấu nguyên âm thì số bàn phím cần gõ thêm sẽ tăng lên tới 153-155%.

Xem xét sơ bộ một bài báo 2135 từ có khoảng 673 dấu mũ, 474 dấu râu và 63 dấu mũ ngược, trong số 12623 byte (hoặc gần như 6 kí tự một từ). Vậy nếu dấu nguyên âm được gõ bằng phím riêng biệt thì số lần gõ thêm sẽ chiếm 10,60%. Mặc dầu 54% số từ là có dấu nguyên âm cũng chỉ chiếm hơn 10,60% số byte trong cả bài viết.

Kiến nghị về cách tổ chức bàn phím chữ Việt

Lựa chọn định nghĩa bàn phím mới hay làm quá tải bàn phím cũ là một lựa chọn đầu tiên cần phải tính tới. Từ các trình bày ở trên chúng tôi nghiêng về hướng làm quá tải các bàn phím đã có sẵn.

Với các bàn phím hiện tại, lựa chọn tiếp theo là nên dựa trên cấu hình của bàn phím QWERTY hay AZERTY. Do tính phổ biến của bàn phím QWERTY hiện tại, câu trả lời nghiêng về QWERTY. Tuy nhiên để vẫn có thể sử dụng bàn phím AZERTY trong các trường hợp cần thiết, các chương trình bàn phím nên cho phép người sử dụng tự định nghĩa lại tương ứng giữa các phím logic và vật lí.

Trong số 12 phím cần thêm vào cho chữ Việt, có 7 phím mô tả cho các nguyên âm thuần Việt, cần được cài đặt ở hai mức tương ứng với các chữ thường và hoa, còn 5 phím chỉ cần cài đặt ở một mức tương ứng với 5 dấu thanh. Có thể chọn 5 vị trí trên hàng chữ số để bố trí cho 5 dấu thanh này.

Các chữ cái tiếng Anh f, j, w, z là những chữ hầu như không có trong các văn bản Việt nên có thể xem xét sử dụng các vị trí đó cho các nguyên âm thuần Việt. Mặt khác, cũng có thể xem xét sử dụng vị trí của các dấu [,], = cho các nguyên âm thuần Việt. Sau đây là một số dự kiến về cách bố trí bàn phím chữ Việt:

Mức 1:

1 2 đ\$ huyền hỏi ngã sắc nặng 9 0 - đ
 q â e r t y u i o p u ơ
 a s d ê g h ô k l ; ' ,
 ă x c v b n m , . /

Mức 2 (Shift):

! @ # \$ % ^ & * () - Đ
 Q Â E R T Y U I O P U O
 A S D Ê G H Ô K L : " ' < > ?

Mức 3: (Alt/Option)

1 2 3 4 5 6 7 8 9 0 =
 w []
 f j
 z

Mức 4 (Shift + Alt / Option)

W + { }
 Z F J

Một dự kiến khác về cách bố trí bàn phím chữ Việt (để tránh xung đột với việc dùng Alt-F, W, J, Z, trong windows)

Mức 1:

ă â ê huyền hỏi ngã sắc nặng ô đ - đ\$
 q w e r t y u i o p u ơ
 a s d f g h j k l ; ' ,
 z x c v b n m , . /

Mức 2 (Shift):

Ă Â Ê \$ % ^ & * Ô Đ
 + Q W E R T Y U I O P U O
 A S D F G H J K L : " ' < > ?
 Z X C V B N M

Mức 3: (Alt/Option)

1 2 3 4 5 6 7 8 9 0 =
 []

Mức 4 (Shift + Alt / Option)

! @ # \$ % ^ & * () { } +

Tài liệu tham khảo

1. ISO 9995-7
2. ECMA 115
3. TCVN 5712
4. VietRest, Seatic
5. BKED 6.0
6. Viet-Std: VISCII & VIQR
7. Ngo Thanh Nhan, *Ve tu bo dau chu Viet*, 1984

Abstract

Computer keyboard for Vietnamese

Vietnamese keyboard is an indispensable tool to use Vietnamese on computers. This paper is intended to present some principals should be considered when designing keyboard for Vietnamese and some problems of its instalation. Vietnamese kyeboard should be conformed main points of views developement of information technology in Vietnam and satisfied four requirements: completeness, consistency, compatibility and simplicity. Based on principals of international kyeboards, according to standards ISO 9995 and ECMA 115, and some Vietnamese particulars, methods of encoding and rendering Vietnamese characters from international kyeboards are considered and commented. Two solutions for designing Vietnamese keyboard satisfying the four above requirements are proposed.