**Vương Quang Khải**

Unicode, Dựng Sẵn hay Tổ Hợp

Chào mừng các bạn đón đọc đầu sách từ dự án sách cho thiết bị di động

*Nguồn:* [*http://vnthuquan.net/*](http://vnthuquan.net/)
Tạo ebook: Nguyễn Kim Vỹ.

**MỤC LỤC**

[Unicode, Dựng Sẵn hay Tổ Hợp](%22%20%5Cl%20%22bm2)

**Vương Quang Khải**

Unicode, Dựng Sẵn hay Tổ Hợp

**Các vấn đề đang tranh cãi**
Hiện nay, có 2 cách để thể hiện (mã hoá) chữ viết tiếng Việt trong bảng mã Unicode:
- Dựng sẵn: coi tập hợp chữ + dấu mũ + dấu thanh như một ký tự duy nhất:
Chữ “Việt” gồm 4 ký tự V, i, ệ, t
- Tổ hợp: coi tập hợp chữ + dấu mũ như một ký tự duy nhất, dấu thanh được tính là một ký tự riêng
Chữ “Việt” gồm 5 ký tự V, i, ê, t và dấu nặng
Vietkey và đa số các chuyên gia về xử lý tiếng Việt trong nước ủng hộ dạng mã hoá dựng sẵn.
Dạng mã tổ hợp được ủng hộ bởi Microsoft và một số công ty trong nước như VASC, Lạc Việt.
Ngoài ra, còn một luồng ý kiến thứ ba cho rằng nên hỗ trợ cả hai dạng mã hoá (chuyên gia của viện CNTT). Ngay trong bảng mã TCVN 6909 cũng chỉ yêu cầu sử dụng Unicode mà không chỉ định rõ nên sử dụng dạng mã hoá nào. Thực ra, ý kiến thứ ba này không hợp lý, vì việc mã hoá không nhất quán sẽ gây nhiều khó khăn cho cả người dùng lẫn nhà sản xuất, làm tăng thêm chi phí một cách không cần thiết, đồng thời gây gánh năng về bảo trì sau này.

**Một số sự thật**
Mã dựng sẵn hiển thị và in ấn tốt với hầu hết các dòng Windows và Office mà không cần phải cài đặt thêm phần cứng, phần mềm, hay hỗ trợ của hệ điều hành. Tuy nhiên, việc xử lý mã dựng sẵn (sắp xếp, chuyển đổi chữ hoa/chữ thường, …) không được hỗ trợ bởi hệ điều hành và các ứng dụng chuẩn. Cần phải cài đặt các thư viện phụ do các hãng thứ ba cung cấp (VD: Vietkey Office)
Mã tổ hợp thường hiển thị rất xấu hoặc bị lỗi trên các dòng Windows, Office cũ. Ngay cả với Windows XP, Office XP cũng vẫn bị lỗi hiển thị (chương trình email Outlook Express) nếu chưa cài thêm phần hỗ trợ tiếng Việt. Khả năng hỗ trợ tiếng Việt không có sẵn mà cũng vẫn phải cài đặt thêm thư viện hỗ trợ, chỉ có điều đây là thư viện của Microsoft.
Do khả năng tương thích tốt của mình, mã dựng sẵn đang chiếm ưu thế tuyệt đối ở Việt Nam. Có thể nói là gần như 100% website hỗ trợ Unicode đều dùng mã dựng sẵn. Nếu thuyết phục các webmaster đổi website sang sử dụng mã tổ hợp, cũng có nghĩa là thuyết phục họ bỏ đi một lượng khách hàng tiềm năng.
Trước đây, Microsoft hỗ trợ tiếng Việt 8 bit theo kiểu tổ hợp. Vì thế hiện nay, hãng quyết tâm đi theo con đường tổ hợp để tái sử dụng các thư viện hỗ trợ tiếng Việt cũ, mà không phải đầu tư sửa lại mã lệnh. Các thư viện cũ đã chạy khá ổn định, việc sửa lại mã lệnh cũng có thể gây ra những lỗi mới.

**Một số ý kiến cá nhân của tác giả**
Cả về mặt kỹ thuật lẫn thực tế, mã dựng sẵn có rất nhiều ưu điểm so với mã tổ hợp (xem thêm các bài phân tích của tác giả Đặng Minh Tuấn trên PC Word A & B)
Phương pháp mã hoá tổ hợp thường chỉ dùng cho các loại chữ viết dạng khối (block) như Trung, Thái, Ả-rập, … Hầu hết các ngôn ngữ thuộc họ Latin đều sử dụng phương pháp mã hoá dựng sẵn. Việc dùng mã tổ hợp cho tiếng Việt chỉ là một cách đi vòng (walkaround) do trước kia, các bảng mã 8-bit không thể biểu diễn được hết các ký tự tiếng Việt dựng sẵn. Với bảng mã Unicode, giới hạn nói trên đã được giải quyết. Microsoft không nên ngại thay đổi mà giữ nguyên cách làm đã lỗi thời.
Microsoft không phải là toàn bộ thế giới điện toán. Chúng ta cần tham khảo thêm hỗ trợ tiếng Việt của các hãng khác. Chẳng hạn như chúng ta đang có chiến lược chuyển dần sang sử dụng Linux, tránh phụ thuộc quá nhiều vào Microsoft. Chúng ta sẽ làm gì nếu cộng đồng Linux ủng hộ mã dựng sẵn?
Sẽ là tốt nhất nếu chúng ta có thể gây áp lực thuyết phục Microsoft chuyển sang sử dụng mã dựng sẵn. Thế nhưng nếu họ khăng khăng không chịu thì sao? Đứng dưới cả góc độ người sử dụng và góc độ nhà sản xuất phần mềm, tôi nghĩ rằng chúng ta vẫn nên đi theo con đường của Microsoft, dù con đường đó có thể còn thiếu hợp lý và không khoa học. Đơn giản là vì con đường đó sẽ giúp chúng ta phát triển và ứng dụng phần mềm nhanh hơn, tiện hơn và rẻ hơn.

**Một số thông tin hành lang**
Chính phủ, cụ thể là bộ Khoa học và Công nghệ, rất không thích phụ thuộc vào Microsoft. Khả năng mã dựng sẵn trở thành chuẩn quốc gia là rất lớn.
Bản thân nội bộ Microsoft cũng đã từng cân nhắc việc chuyển sang hỗ trợ mã dựng sẵn. Có điều nếu làm như vậy thì công sức đầu tư làm việc với tổ hợp từ trước đến giờ sẽ phải bỏ đi hết. Chính vì vậy mà Microsoft quyết tâm ủng hộ mã tổ hợp, và gần như chắc chắn sẽ không thay đổi ý kiến.
**Dự đoán kết quả**
Có 3 kịch bản sẽ xảy ra trong tương lai:
- Mã dựng sẵn được sử dụng rộng rãi, mã tổ hợp bị lãng quên, giống như bảng mã tiếng Việt CP1258 của Microsoft không có ai sử dụng.
Khả năng: 60%
- Mã dựng sẵn trở thành chuẩn lý thuyết, mã tổ hợp trở thành chuẩn thực tế tự phát, giống như trường hợp miền Nam vẫn ưa sử dụng font VNI dù chính phủ đã quyết định chọn ABC là font chuẩn.
Khả năng: 25%
- Chính phủ và Microsoft cùng phối hợp hỗ trợ một bảng mã duy nhất.
Khả năng: 15%
Mặc dù khả năng thứ ba là thấp nhất, nó sẽ là kịch bản tốt nhất cho tất cả mọi người: người sử dụng Việt Nam, nhà sản xuất phần mềm Việt Nam, chính phủ Việt Nam, và Microsoft.

Lời cuối: Cám ơn bạn đã theo dõi hết cuốn truyện.
Nguồn: http://vnthuquan.net
Phát hành: Nguyễn Kim Vỹ.

Nguồn: Thái Nhi
Được bạn: Thành Viên VNthuquan đưa lên
vào ngày: 27 tháng 12 năm 2003