

Phương hướng mới cho Tự điển chữ Nôm (Online Dictionary of Nôm Characters, a New Approach)

Nguyễn Hữu Vinh

(also known as Yuan Yu-Rong)

Quality Control Manager, Industrial Technology Research Institute, Hsin Chu, Taiwan

Abstract

With the current computer and Internet technologies, it is feasible to compose a clear, precise, complete, easy-to-use electronic Nôm characters dictionary that opens the door to new researching capabilities. The method of collecting Nôm characters by searching through Nôm literature is recommended as a useful and convincing solution. Each Nôm character in the dictionary needs to be accompanied with quoted excerpts from multiple Nôm writings existed from old time. The dictionary also needs to explain the composing structure of each character in the context of illustrated examples. An electronic Nôm dictionary needs to incorporate current modern technologies which make it easy to broadcast through the Internet and easy to search (by Quoc Ngu, character, Unicode number, or by a set of characters, ...). Most importantly, it has to be compatible among computers and computer platforms. In order to satisfy these demands, a Nôm font conforming to Unicode standards needs to be created. In addition, a database structure needs to be defined from the beginning with essential fields such as characters, Unicode number, Quoc Ngu pinyin, structure of the characters, examples, reference sources, etc. Two user interfaces are required: an end-user's interface and a developer's interface. The end-user's interface provides searching and querying capabilities to retrieve data from the database. The developer's interface allows the Working Committee to record data and to collaborate all aspects of the work. Besides for the normal searching capabilities, the dictionary can also be expanded to provide statistical and translating tools. The Committee with different levels of knowledge and experiences not only in Nôm characters and literature but also in computer technology, from different parts of the world, will bring more complete and comprehensive results, which are much more difficult to achieve by an individual. To ensure consistency and quality of the dictionary, the Committee needs to establish a working process, which conforms to proven project management, software development and quality assurance principles. Lastly, we will share some of design specifications of recently released Online Dictionary of Nôm Characters with Excerpts.

Tóm lược

Với các phương tiện điện toán và Internet hiện có, việc biên soạn một quyển tự điển chữ Nôm điện tử rõ ràng, chính xác, đầy đủ, dễ dùng và có thể mở ra những chiều hướng nghiên cứu mới là điều rất khả thi. Phương thức thu nhập dữ kiện bằng cách lật soát các hình thái chữ Nôm từ những văn bản Nôm được xem như là một giải pháp hữu ích và có tính thuyết phục. Mỗi một chữ Nôm trong tự điển đều phải được dẫn chứng bằng các thí dụ cụ thể trích dẫn từ các văn bản Nôm khác nhau. Tự điển còn cần phải có phương cách giải thích cấu trúc của chữ Nôm theo ngữ nghĩa của chữ. Một tự điển chữ Nôm điện tử cần phải được hỗ trợ bởi những phương tiện kỹ thuật hiện đại như dễ truyền thông qua mạng Internet, dễ tra cứu (bằng âm Quốc Ngữ, mặt chữ, mã chữ, hoặc tập hợp chữ, v.v), và nhất là có tính cách quốc tế, máy nào cũng dùng được. Để đáp ứng được các nhu cầu đó, cần có một bộ font chữ Nôm theo mã chuẩn quốc tế (Unicode). Thêm vào đó, cần có một cơ sở dữ kiện (database) được xác định ngay từ bước đầu để chứa đựng

các thông tin về một chữ Nôm như mặt chữ, mã chữ, âm Hán Việt, âm đọc, cấu trúc, văn liệu, v.v... Kế đến là một giao diện cho người dùng (end-user's interface), cho phép người dùng tra cứu và tìm kiếm (search) các thông tin chứa đựng trong kho dữ kiện này một cách dễ dàng và nhanh chóng. Ngoài ra, cần một Giao diện (interface) dành riêng cho Ban Biên Tập (BBT) để nhập dữ kiện vào kho dữ kiện và thảo luận những vấn đề liên quan đến mọi khía cạnh của tự điển. Ngoài các chức năng tra cứu của một tự điển thông thường, tự điển còn có thể mở rộng để hỗ trợ công việc thống kê và phiên âm. Một BBT với nhiều khả năng và trình độ khác nhau về cả chữ Nôm lẫn kỹ thuật điện toán, cư ngụ ở nhiều nơi trên thế giới, sẽ mang lại những thành quả mà một cá nhân khó có thể làm nổi. BBT cần phải xác định các quy trình làm việc rõ ràng, với một số nguyên tắc về quản lý dự án (project management), phát triển phần mềm (software development) và quản lý chất lượng (software quality assurance) để bảo đảm tính thống nhất và chất lượng của tự điển. Cuối cùng là vài khía cạnh của quyển Tự Điển Chữ Nôm Trích Dẫn được thiết kế theo chiều hướng trên.

Tại sao Tự Điển Nôm trên mạng

Một tự điển được xem là thực dụng và hữu ích tối thiểu phải hội đủ những yếu tố: rõ ràng, chính xác, đầy đủ, dễ dùng và hiệu quả. Đối với một tự điển chữ Hán thì việc sắp xếp tra tìm chữ là một trong những yếu tố ưu tiên hàng đầu để phù hợp với những yếu tố nêu trên. Trong một tự điển chữ Nôm, thì các yếu tố trên lại có phần phức tạp hơn nữa. Chữ Nôm không được điển chế, việc xác định hình và nghĩa của chữ chưa được rõ ràng, cho nên công việc làm nên một tự điển chữ Nôm gặp những trở ngại cần phải giải quyết. Chữ Nôm nên được sắp xếp thế nào cho tiện lợi, tra cứu cho nhanh chóng, nên có thêm những dữ kiện cần thiết nào, làm sao ghi lại được rõ ràng và chính xác hình và nghĩa, dùng trong trường nào, thuộc văn cảnh nào, tác phẩm nào, tại sao đọc như thế này thế nọ v.v Rõ ràng, một số vấn đề như trên cần phải được đề ý và giải quyết thích đáng. Với sự áp dụng kỹ thuật máy điện toán vào công việc bảo tồn chữ Nôm ngày càng trở nên thực tế, thì một cuốn tự điển chữ Nôm online trên mạng Internet có khả năng giải quyết một số lớn vấn đề nan giải nêu trên.

Mã chữ nào và loại phông nào thích hợp

Từ nhiều năm nay, một số công ty nước ngoài như Công ty Dynacomware của Đài Loan đã thiết kế một loại phông đặc biệt cho chữ Nôm, tuy nhiên nhưng không biết vì lý do gì phông này không thấy cho lưu hành trên thị trường. Riêng Mojikyo Institute ở Japan thì đã cho ra thị trường phần mềm nổi tiếng Kim Tích Văn Tự Kính (今昔文字鏡単漢字 10 万字版) chạy dưới Windows 2000 hoặc Japanese version XP, nhưng chỉ có khoảng 2000 chữ Nôm, với lại phần mềm này không sử dụng được với Unicode, cho nên không thực dụng đối với người Việt. Ngày nay, một số lớn chữ Nôm đã được chuẩn hoá và cấp mã Unicode nên sự xúc tiến việc chế tạo các phông chữ Nôm dạng TrueType để ứng dụng cho các phần mềm máy điện toán đã trở nên quan trọng, cần thiết và cấp bách. Chọn một loại phông thích hợp để phát triển các phần mềm ứng dụng chữ Nôm phải là điều suy nghĩ đầu tiên của các nhà thiết kế phần mềm. Trước khi chữ Nôm được mã hóa vào Unicode, thì trong nước cũng đã có người xúc tiến việc chế tạo phông, song vì chúng ta chưa có được một bộ mã quốc gia, do đó việc chế tạo phông rơi vào tình trạng bế tắc. Ngay cả sau khi chữ Nôm được đưa vào Unicode, các phần mềm ứng dụng cũng chưa để ý tới. Một phần bởi lẽ các phần mềm dùng để làm data processing như Microsoft Word cũng chưa quan tâm, một phần bởi lẽ hiển thị chữ Unicode trong các vùng CJK Unified Ideographs của Unicode đòi hỏi kỹ thuật hiện đại và tiên tiến nhất mới thực hiện được. Ví dụ, để hiển thị chữ trong vùng CJK Unified Ideographs B, Microsoft đưa ra kỹ thuật surrogate để xử lý, vừa phức tạp vừa đòi hỏi máy tính phải xử lý được đa ngôn ngữ và chạy với vận tốc nhanh. Với điều kiện kỹ thuật và kinh nghiệm về phông chữ 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 trong khi việc phát triển các phần mềm ứng dụng Unicode trên thế giới đang còn nằm trong thời kỳ phôi thai. Với lại, số chữ Nôm được cấp mã Unicode dù đã được khoảng 10,000 chữ, song con số này vẫn chỉ được xem như là rất khiêm tốn. Còn rất nhiều chữ Nôm chưa có mã Unicode nằm nhan nhản trong các tác phẩm Nôm. Chỉ hai việc nêu trên, cũng đủ làm cho việc phát triển các phần mềm căn bản cho việc ứng dụng chữ Nôm như Bàn gõ, Tự điển, database, webpage v.v đều bị trở ngại lớn. Để dùng được chữ Nôm với bộ mã Unicode, một điểm kỹ thuật cần xử lý là các chữ Nôm chưa có mã Unicode phải đưa vào vùng mã dùng riêng (Private Use Area) của Unicode. Song mã trong vùng này thường được các phần mềm khác sử dụng, nên hay xảy ra xung đột lẫn nhau. Muốn tránh trở ngại này, thì có thể sử dụng các vùng Private khác, song lại bị trở ngại kỹ thuật surrogate như đã nói ở trên. Việc thiết kế một bộ phông thực dụng cho chữ Nôm là điều không phải dễ dàng. Chữ Nôm vì bị xếp nằm rải rác trong các vùng CJK Unified Ideographs (4E00-9FAF), CJK Unified Ideographs Extension A (3400-4DBF), CJK Unified Ideographs Extension B (20000-2A6DF) và sau này sẽ là vùng CJK Unified Ideographs Extension C, do đó nếu một bộ phông chữ Nôm chỉ lọc ra các chữ Nôm có trong các vùng chữ mà loại bỏ các chữ Hán trong vùng thì sẽ giới hạn tính chất thực tiễn của phông khi dùng chung với chữ Hán trong cùng một ứng dụng (application). Nhưng nếu bao gồm toàn bộ chữ Hán có trong các vùng đó thì bộ phông trở nên to lớn, kênh càng và nhất là đòi hỏi thời gian và công sức rất là to lớn để thiết kế nên bộ phông đó! May rằng, chúng ta đã có một bước tiến đột phá là việc ra đời của hai bộ truetype phông Hán Nôm A và Hán Nôm B để hiển thị chữ Nôm và chữ Hán mã Unicode do Chân Nguyên Đỗ Quốc Bảo (Germany), Tô Minh Tâm (USA), Ni sinh Thiên Viện Viên Chiếu (Việt Nam) và các tác giả của bộ Tự Điển Chữ Nôm Trích Dẫn online cùng nhau thực hiện. Bộ phông Han Nom A và B xử lý chữ Nôm rất tiên tiến về mặt kỹ thuật cho đến số lượng chữ vô cùng lớn nằm trong toàn thể các vùng CJK Unified Ideographs, mà sau này có thể dùng được luôn cho Extension C! Ngay cả China và Taiwan, cho tới bây giờ vẫn chưa có 1 bộ phông như bộ phông Han Nom A/B chúng ta đang có. Ở Japan, tổ chức Mojikyo Institute vẫn chưa có bộ phông đạt tới phạm vi lớn trong Unicode như bộ phông Han Nom A và B này. Với điều kiện kỹ thuật và kinh nghiệm về chữ còn rất hạn chế ở nước ta hiện nay, việc thực hiện được công việc đồ sộ và phức tạp như vậy là điều không phải dễ dàng và đáng khích lệ. Với 2 bộ phông lớn này, chúng ta đã có thể xử lý và hiển thị được tất cả các chữ Hán và chữ Nôm có mã trong các vùng chữ của Unicode. Ngoài ra, bộ phông này cũng đã mã hóa tạm thời khoảng hơn 2000 chữ Nôm chưa có mã Unicode và để vào 2 vùng Private Use bắt đầu từ U+E000 và U+F0000.

Tóm lại, sự phát triển phần mềm ứng dụng chữ Nôm có khả thi và thực dụng hay không đều hoàn toàn phụ thuộc vào loại phông và mã chữ chọn dùng cho phần mềm đó, những yếu tố cần để ý tới như sau: 1) Mã Unicode hay mã quốc gia hay là riêng của một phần mềm, của tổ chức nào đó, 2) Nếu là Unicode, có nên phát triển bộ phông trong đó bao gồm các loại chữ vuông khác cùng nằm trong cùng một vùng chữ, 3) Nên có bao nhiêu chữ chưa có mã Unicode nhưng được mã hóa tạm thời? và 4) Chọn vùng Private Use ở đâu?, nếu đặt ngoài vùng U+E000 thì phải có phương cách nào để hiển thị chúng vì nếu dùng Iexplorer thì kỹ thuật Surrogate bây giờ giới hạn hiển thị chữ trong vùng B mà thôi.

Computer platforms quan trọng hay không?

Nói tới Tự Điển Online thì người ta nghĩ ngay tới một trang web trong đó người dùng có thể tra cứu ngay chữ đó. Thông thường, các webpages đều được viết bằng dạng Hypertext Markup Language (html), nhưng html có những bất lợi như là phải online, vào web mới xem được, và người khác có thể lấy về dễ dàng, lấy nguyên cả website về làm data riêng cho họ! Nếu để ý tới các phần mềm bổ túc khác ngoài việc thiết kế bằng dạng Hypertext Markup Language, thì một tự điển online cũng phải nên chú trọng đến các yếu tố khác, như dễ di chuyển (migration) vào các

platforms khác, vận tốc nhanh chậm v.v. Một trong những khuynh hướng hiện tại cho application programs là sử dụng ngôn ngữ Java để thiết kế phần mềm. Ngôn ngữ Java với những ưu điểm như là chỉ cần coding 1 lần, không cần thay đổi gì cả nhưng có thể chạy online lẫn offline; và có thể chuyển (migration) vào tất cả các platforms khác như Windows, Macintosh, Linux, Pocket PC, Palm, Cellular phone v.v. Chúng ta không thể chuyển đổi platform linh động như thế mà không dùng tới Java. Ngoài Java ra, chưa có programming language nào có thể làm được chuyên này so với các language khác. Chỉ có java applet mới có thể embed được trong một webpage. Tiện lợi là ở bất cứ nơi đâu, như tại thư viện, đi du lịch vòng quanh thế giới, ở đâu cũng có thể tra cứu tự điển được mà không cần phải có web browsers. Khuyết điểm của Java là chạy hơi chậm, nhưng có thể bù lại bằng cách nâng cấp máy tính lên hàng vận tốc cao hơn, hiện đại hơn, vì dù sao thì xử lý chữ Nôm vẫn đòi hỏi máy móc hiện đại và tiên tiến mới làm được. Mặc dù sử dụng Java programming trong một tự điển online làm cho việc hiển thị chữ có phần chậm, song trong bối cảnh hiện tại, tự điển chữ Nôm online chạy trên nhiều platforms là điều cần thiết cho nhiều người dùng khác nhau.

Một phương pháp tra chữ mới và đặc biệt

Hiện nay, những tự điển có chú thích chữ Nôm có bộ Đại Nam Quốc Âm Tự Vị của Huỳnh Tịnh Của, Dictionarium Anamitico Latinum của AJ. L. Taberd, Dictionnaire Annamite Francais của J. F. M. Génibrel, Dictionnaire Annamite Francaise của Bouet etc. Những sách xưa này chủ yếu là những tự điển tra chữ từ âm Quốc Ngữ, nên các mục chữ đều được sắp xếp theo mẫu tự Latinh. Trước đó, trong khoảng thế kỷ XVI đến XIX còn có các sách được xem như là tự điển Hán Việt ghi tiếng Việt bằng chữ Nôm, như Chi Nam Ngọc Âm Giải Nghĩa thời Lê sơ, Nhật Dụng Thường Đàm của Phan Đình Hồ thời Nguyễn sơ, Tự Học Toàn Yếu của Ngô Thời Nhậm thời chúa Nguyễn, Tự Đức Thánh Chế Tự Học Giải Nghĩa của vua Tự Đức v.v đều căn cứ vào những cách sắp xếp nào đó liên hệ đến chữ Hán, nên muốn tra tìm 1 chữ Nôm đều rất khó khăn. Vào thời cuối thế kỷ XX sau này, còn có Đại Tự Điển Chữ Nôm của Vũ Văn Kính, Bảng Phiên Âm Nôm Việt của Trương Đình Tín, Tự Điển Chữ Nôm của Takeuchi Yonosuke, Từ điển Từ Vựng Lịch Sử Chữ Nôm của Paul Schneider v.v mỗi cuốn đều có những cách sắp xếp chữ theo những hệ thống khác nhau, nhưng tựu chung một vấn đề nan giải của một cuốn tự điển chữ Nôm trên giấy là làm thế nào sắp xếp chữ cho có hệ thống, dễ dùng, dễ tra cứu, nhanh chóng và có hiệu quả. Tự điển chữ Nôm nên được sắp xếp như thế nào? Cách xếp đặt được người ta nghĩ ngay đến là cách xếp đặt theo bộ thủ như các tự điển chữ Hán. Nhưng chữ Nôm lại rất phức tạp, vị trí của phần được xem như là bộ thủ lại không nhất định, khó nhận thấy bộ thủ chủ yếu ở đâu, lấy phần viết bên trái, bên phải, hay là bên trên, bên dưới, hay là lấy phần biểu ý làm bộ thủ đều không có quy định nhất thiết. Một điều bất thành văn để quyết định đâu là bộ thủ cho chữ Nôm là lấy phần biểu ý hay là bộ thủ của phần biểu ý để làm bộ thủ cho toàn chữ. Tuy nhiên, điều này trở thành không đúng sau khi chữ Nôm được đưa vào Unicode. Các chuyên gia Tàu, có thể không hiểu hoặc vì thiên kiến đã cho không cho mã cho những chữ Nôm có hình dạng trùng hợp với chữ Hán. Có những chữ Nôm trùng dạng chữ Hán như chữ Đền 焜, chữ này là 1 chữ Nôm có hình dạng với chữ Hán 焜, Unicode không cấp thêm mã cho chữ, chúng ta phải dùng chung với chữ Hán, và vì vậy chữ Nôm này có bộ thủ là 田 thay vì 火 như chúng ta đã nghĩ như thế. Có những chữ Nôm ý nghĩa giống nhau, phần biểu ý giống nhau nhưng vị trí của nó đối với phần biểu âm khác nhau, Unicode lại xếp vào các bộ thủ khác nhau, thay vì xếp vào bộ thủ của phần biểu ý của các chữ đó, ví dụ: Hai chữ Đuôi 墀, 隹 này được xếp vào hai bộ thủ khác nhau. Chữ Nôm mượn âm Hán Việt của chữ Hán để ghi âm, nên bộ thủ của chữ Hán đó không biểu thị được nghĩa của nhiều âm Nôm của cùng một chữ Nôm trùng dạng chữ Hán đó. Lại còn có những chữ Nôm đặc biệt như 夕(làm), 垂(xông), 仄(ây), 羊(tiêu), 冫(có), 米(vê), 重(xong) thì nên xếp vào bộ

thủ nào cho thích hợp. Như thế, vấn đề nhận định và sắp xếp bộ thủ trong chữ Nôm quả là một là một điều chưa giải quyết được. Do vậy, sắp xếp chữ trong tự điển chữ Nôm theo bộ thủ không phải là điều ổn thỏa nhất. Cách sắp xếp khác là xếp đặt theo số nét. Thực ra thì cách xếp đặt theo số nét cũng phải kết hợp với cách xếp bộ thủ. Còn cách xếp theo tổng số nét cũng không kém phần phức tạp đối với người Việt nhưng theo Trần Văn Giáp thì phương pháp này là giản dị và tiện lợi nhất so với các cách xếp đặt khác. Còn những chữ Nôm trùng với chữ Hán thì còn có thể mượn cách sắp xếp theo cách đọc PinYin, ZhuYin hoặc các phương pháp xếp đặt đặc thù của China hay là của Taiwan v.v. Những cách xếp đặt khác và mới hơn như là xếp theo số mã hoặc theo phương pháp mã bốn góc (tứ giác hiệu mã), cách này cũng khoa học và thực tế, chỉ cần trông chữ và tìm hình ngay, nhưng phải nhớ tới bốn chữ số thì cũng phiền toái lắm mà không phải người Việt nào cũng rành cách nhận mặt chữ Nôm. Hiện tại, cách sắp xếp chữ trong một tự điển chữ Nôm trên giấy thường có các phương cách sắp xếp thông dụng như trong các tự điển chữ Hán: bằng âm Nôm, âm PinYin của China, Zhuyin của Taiwan, bằng phương pháp mã bốn góc, theo số mã Unicode, bằng số nét kết hợp với bộ thủ, bằng tổng số nét v.v. Mỗi phương cách đều có chỗ đứng riêng của nó. Như vậy, đối với phương cách sắp xếp để tra tìm chữ cho một tự điển chữ Nôm online thì sao? Ngoài các phương cách tra tìm nêu trên, tự điển chữ Nôm online cần phải cho phép tìm kiếm (searching) chữ Nôm "theo mặt chữ viết", nghĩa là, căn cứ trên các thành phần cấu tạo nên chữ Nôm đó. Đây phải là một chức năng đặc biệt và phải làm sao để cho chức năng tìm kiếm này của tự điển làm việc một cách thông minh, dễ hiểu, dễ dùng, nhanh chóng và có hiệu quả. Người dùng tự điển, khi gặp một chữ Nôm có thể không biết âm đọc là gì, không biết thuộc bộ thủ nào, không rõ số mã Unicode, nhưng chắc chắn là biết được thành phần cấu tạo nên chữ đó, chỉ cần tìm kiếm (searching) chữ này bằng cách tra theo tập hợp của nhiều thành phần của chữ đó và sẽ tìm ra những chữ Nôm có những thành phần đó có trong tự điển. Lợi dụng đặc tính tìm kiếm (search) nhanh chóng của computer, thì việc sắp xếp chữ Nôm trong database sao cho việc tra tìm chữ được trở nên dễ dàng và thông minh hơn. Database của tự điển cho Nôm online cần phải có một vài Trường (fields) chứa đựng những thông tin (information) cần thiết về tập hợp thành phần cấu tạo nên một chữ Nôm, để sau này lợi dụng một searching algorithm nào đó để tra tìm (searching) chữ Nôm có thành phần cấu tạo đã ghi trong fields của chữ đó trong cơ sở dữ liệu (database). Đối với một tự điển chữ Nôm online thì đây là một phương cách tìm kiếm hiệu quả nhất cho tới nay, nhất là cho người dùng khi tra tìm một chữ Nôm lạ không rõ bộ thủ, không biết âm đọc, không biết mã chữ. Nếu thành phần của chữ Nôm này bao gồm bộ thủ của chữ, thì có thể nói rằng đây là phương pháp tổng quát hóa phương cách tìm bằng bộ thủ nhưng nhanh chóng và có hiệu quả hơn nhiều. Ví dụ: Tra tìm chữ 𠄎, đối với tự điển bằng giấy, có thể vì không biết âm Nôm, tự điển không có phương pháp mã bốn góc, người tra tự điển có thể phải tốn thì giờ tra loại bộ thủ, nhưng tra bộ thủ nào, bằng bộ 日, bằng bộ 月, hay bằng bộ 心, vừa mất thì giờ vừa mệt nhọc! Với thí dụ chữ 𠄎 này, người dùng tự điển Nôm online hoặc là searching bằng 2 keys: 日 và 月 có thể tìm ra các chữ sau: 棚, 棚, 明, 𠄎, 盟, 腊... hoặc là searching bằng 2 keys: 日 và 心 có thể tìm ra các chữ 怛, 愀, 棚, 𠄎, 愀, 慢, 愀... hoặc là searching bằng 2 keys: 心 và 月 có thể tìm ra các chữ 𠄎, 𠄎, 愀, 愀, 愀, 愀... hoặc là searching cả bằng 3 keys: 日, 月, 心 thì có thể tìm ra ngay các chữ 𠄎, 棚. Tiện lợi của cách tra theo tập hợp thành phần của chữ này là nhanh chóng và hiệu quả vì người dùng có thể dựa trên mặt chữ để tra tìm khỏi phải cần biết bất kỳ dữ kiện nào khác như bộ thủ, mã Unicode, âm đọc v.v.

Tóm lại, ngoài các cách sắp xếp chữ Nôm theo các phương cách đã được dùng trong các Tự điển chữ Hán và Nôm trên giấy, một tự điển chữ Nôm online cần phải có thêm những phương cách nhận diện và tra tìm chữ bằng cách cho phép người dùng tìm kiếm (searching) bằng thành

phần tạo nên chữ đó. Cho đến nay, đây là một phương cách được xem như là có hiệu quả nhất để tra tìm chữ trong một tự điển chữ Nôm online .

Kết cấu chữ Nôm

Những phép cấu tạo chữ của chữ Nôm đã vượt qua khỏi phép Lục thư của chữ Hán đã nói lên sự cố gắng tạo chữ đặc thù khác với phương pháp dùng cho chữ Hán, mang một tinh thần dân tộc riêng chứ không nhất nhất đều phải rập theo khuôn khổ chữ Hán. Trước thập niên 70, những nhà nghiên cứu nổi tiếng như Dương Quảng Hàm, Hoàng Xuân Hãn, Vương Lực, Chen Ching Ho, Văn Hựu, Trần Văn Giáp, Đào Duy Anh, Bửu Cầm, Vũ Văn Kính v.v đều cho rằng chữ Nôm đã sử dụng 3 phương pháp tạo chữ trong 6 phương pháp (Lục Thư) tạo chữ của chữ Hán, đó là ba phép Giả tá, Hình thanh và Hội ý. Chen Chin Ho năm 1949 cũng đã phân loại cấu trúc chữ Nôm theo ba loại là Giả tá, Hình thanh và Hội ý. Trong đó Chen cho rằng có ba loại Giả tá, hai loại Hình thanh và đưa ra năm chữ Hội ý. Trước đó, Dương Quảng Hàm năm 1942 cho rằng có năm loại Giả tá, hai loại Hình thanh và đưa ra một chữ Hội ý. Sau này, Đào Duy Anh năm 1975 giải thích chi tiết hơn năm loại Giả tá, hai loại Hình thanh và chỉ đưa ra sáu chữ Hội ý. Các cách phân chia trên có thể nói là đã bao gồm hầu hết các loại chữ Nôm. Tuy nhiên trên thực tế, còn có nhiều loại chữ Nôm có cách cấu trúc đặc biệt thì cả ba phép tạo chữ này đòi hỏi linh động hơn khi giải thích sự cấu tạo của chữ. Ví dụ như loại chữ Nôm đặc biệt mang âm Việt cổ, hoặc chữ bị viết giản nét hoặc dùng một chữ Nôm khác làm thành tố biểu âm thì ba phương pháp này hơi gặp lúng túng trong sự giải thích sự cấu tạo chữ. Sau thập niên 70, các nhà nghiên cứu Nôm nhận thấy những mô hình cấu tạo chữ Nôm thời trước đã nêu vẫn chưa được hoàn hảo để giải thích nhiều trường hợp đặc biệt, nên đã đi sâu hơn như là tìm hiểu sự liên quan giữa âm Nôm của toàn chữ với âm Hán Việt của thành tố biểu thị âm đọc, hay là dựa trên những phương pháp ngữ âm học lịch sử, hoặc là chia chữ Nôm ra thành loại “có âm đọc” và “không có âm đọc”. Trước hết, Nguyễn Tài Cẩn năm 1976, đưa ra mô hình cấu trúc chia chữ Nôm ra bảy loại dựa trên căn bản chữ “mượn Hán” gồm chữ Hán Việt, gốc Hán Việt, viết tắt, giả tá và loại chữ “Sáng tạo” trong đó gồm chữ hình thanh, hội ý, viết tắt, dấu nháy và âm Việt cổ. Công trình nghiên cứu và mô hình của Nguyễn Tài Cẩn dẫn đến các công trình nghiên cứu khác sau này với những lối nhìn khác hơn và chi tiết hơn. Lê Văn Quán năm 1981 đưa ra mô hình giải thích mới trong đó chữ Nôm được chia ra thành 14 loại, trong đó 6 loại thuộc nhóm chữ dựa trên chữ Hán hiện có, còn 8 loại sau thuộc nhóm chữ “Sáng tạo” gồm chữ hình thanh, hội ý, viết tắt, dấu nháy và âm Việt cổ. Sau này Nguyễn Ngọc San dựa vào âm đọc chia chữ Nôm ra 14 loại trong đó chỉ có 4 loại là trực tiếp mượn âm Hán Việt. Với ba mô hình giải thích cấu tạo chữ Nôm của ba học giả nói trên cho thấy sự đa dạng của cấu trúc chữ Nôm không những chỉ đơn thuần gồm ba loại Giả tá, Hình thanh và Hội ý căn bản thường thấy mà còn nhiều loại chữ có cấu trúc đặc biệt, nói lên sự đa dạng và phong phú của âm Nôm hơn âm tiếng Hán. Dù các nhà nghiên cứu chữ Nôm đã đưa ra nhiều lối phân loại chữ Nôm khá hoàn chỉnh, nhưng cho đến bây giờ vẫn chưa có được một mô hình kết cấu (Classification schemes) thật hoàn hảo để giải thích sự kết cấu của chữ Nôm, nhất là vấn đề kết cấu của những chữ Nôm có tính chất đặc biệt như 1) Chữ Nôm âm Việt cổ sử dụng những ký tự chữ Hán như 巴, 巨, 車, 一, 多, 司, 波, 阿, 坡, 丿, 麻, 破, 可, 羅 v.v để ghi âm tiết phụ. Những ký tự này có lúc đi cùng trong một chữ vuông khác như chữ Về 𪗇 hoặc là dùng 2 chữ Hán riêng biệt để ghi một âm Nôm như trong Phật Thuyết Đại Báo Phụ Mẫu Ân Trọng Kinh, ví dụ, dùng hai chữ 多迷 đi cùng nhau để ghi một âm Về, hoặc dùng hai chữ 麻例 đi cùng nhau để ghi một âm Lờ, hoặc dùng hai chữ 𠂇打 đi cùng nhau để ghi một âm Đá. 2) Chữ giả tá Nôm, là loại chữ Nôm được cấu tạo bởi một chữ Nôm khác cùng âm hay tương tự, song ý nghĩa của nó khác hẳn với ý nghĩa với chữ Nôm mượn làm chữ biểu âm, 3) Chữ Nôm từ chữ Hán viết giản nét, thí dụ điển hình nhất là chữ Làm 𠂇 mà phần nhiều các nhà Nôm học cho rằng nó là cách viết

giản nét chữ 爲, chữ Phút 丿 là giản nét chữ Phân 分, 4) Chữ Nôm do chữ hình thanh viết giản nét, có thể vì chữ hình thanh này có quá nhiều nét, khó nhớ, khó viết nên trong quá trình hình thành chữ viết một phần của chữ đã bị cố ý giản lược đi. Ví dụ: Đi 移, là chữ hình thanh, tập hợp của 去移, nhưng chữ 禾 đã bị lược bỏ; chữ Theo 躑 là tập hợp của bộ Túc với chữ Thiêu 燒 giản nét bộ Hỏa; chữ Nôi 淫 là tập hợp bộ Thủy với chữ 𩚑 giản nét bộ Thực 食, 5) Chữ Nôm do ảnh hưởng của chữ cùng đi đôi, loại này thông thường mang bộ thủ giống bộ thủ của chữ thường hay đi đôi với chữ. Ví dụ: Chữ Gân 𠵽 thường hay đi đôi với Xa 𠵽, (xa gân) nên Gân có bộ thủ là Bôi, giống như bộ thủ của Xa, trong khi bộ thủ Bôi trong chữ Gân không giúp thêm ý nghĩa gì cho Gân cả. Do vậy, để lý giải tại sao chữ Nôm được viết như thế này như thế kia, điều này cũng góp phần không nhỏ trong việc giúp người sử dụng nhớ mặt chữ, nhận định ra âm đọc, tự điển chữ Nôm cần phải có phương cách giải thích sự kết cấu của chữ Nôm để xem chữ Nôm này trong trường hợp này đọc như thế này, trường hợp kia giải thích ra làm sao cho thích hợp với văn mạch.

Dẫn chứng và sao lục

Để biên soạn một bộ Tự Điển Chữ Nôm, các nhà làm Tự Điển thông thường mong muốn Tự Điển của mình có một số lượng chữ Nôm đồ sộ, nên công việc trước tiên là góp nhặt, trích dẫn tất cả các chữ Nôm trong các tự điển đã xuất bản. Song làm như thế cũng chưa đủ, vì các tác phẩm trên, không được biên soạn trên cơ sở liệt soát tất cả các tác phẩm chữ Nôm đương thời, cho nên có nhiều chữ Nôm thời trước đó không được ghi vào. Ví dụ chữ Gió 𩚑, 愈, 𩚑, 𩚑, 𩚑, 𩚑, 𩚑, 𩚑 thì các tự điển chữ Nôm đã nêu trên cũng chỉ có một vài chữ. Nhiều lúc lại có những chữ Nôm chỉ thấy ghi trong tự điển mà không thấy dùng trong các tác phẩm Nôm. Nếu một tự điển Nôm chỉ chú trọng tới việc làm sao thu nhặt được thật nhiều chữ để cho số lượng chữ được đồ sộ mà không chú tâm tới vấn đề chữ Nôm này có thật hay không, đã xuất hiện trong tác phẩm Nôm cổ nào, cách dùng trong văn cảnh nào, thì giá trị thực dụng của cuốn tự điển này bị giảm đi rất nhiều. Do việc sưu tầm cứ liệu chữ Nôm còn hạn chế và chưa có hệ thống, chưa nói đến việc bao quát toàn bộ, thế nên theo Trần Văn Giáp thì có những trường hợp chữ Nôm được “nghĩ ra”, “suy ra” hoặc “sáng chế ra” hoặc không thấy trong văn bản nào cả. Do vậy, tự điển chữ Nôm phải trung thành với văn bản, phải ghi lại rõ ràng xuất xứ, phải có cách giúp lý giải tính thích hợp trong văn mạch và giải thích kết cấu (classification schemes) của chữ Nôm đó trong những trường hợp dùng trong một tác phẩm nào đó. Nhờ tính chất dung nạp dữ kiện rộng lớn của computer, một tự điển chữ Nôm online có thể thu lượm và đúc kết càng nhiều càng tốt cách dùng của một chữ Nôm qua nhiều thí dụ ứng dụng chữ này trong nhiều trường hợp trong các văn bản khác nhau. Muốn cho tự điển Nôm thâm thập được một cách đầy đủ các chữ Nôm thông dụng và đặc biệt có giá trị thì cần phải liệt soát tất cả các tác phẩm chữ Nôm có trong dân gian. Dĩ nhiên, chúng ta không thể liệt soát được hết tất cả các tác phẩm Nôm, nhưng ít nhất là phải liệt soát những tác phẩm nổi tiếng, quan trọng và tiêu biểu của nhiều thể loại, thuộc nhiều thời kỳ trong lịch sử; cũng lựa chọn những dị bản tiêu biểu của một tác phẩm; cũng như đưa vào tự điển những văn bản Nôm có âm hưởng địa phương khác nhau của đất nước. Mỗi chữ Nôm trong tự điển, với một hay nhiều âm đọc, với một hay nhiều nghĩa, đều phải được dẫn chứng bằng các thí dụ cụ thể trích dẫn từ các tài liệu, văn bản Nôm khác nhau và được ghi chú xuất xứ chính xác, nếu có, gồm tên tác phẩm, thời điểm xuất bản của tác phẩm, tên nhà xuất bản, trang/tờ, câu v.v. Bằng phương cách này, tự điển chữ Nôm đó sẽ có được những chữ/từ thuần Việt cổ, sẽ tìm ra được rất nhiều âm đọc cho một chữ Nôm, ví dụ, chữ 朗 chúng ta sẽ tìm ra được rất nhiều âm đọc trong nhiều văn cảnh khác nhau: lảng, lãng, lẫng, lẫng, lặng, lặng, lũng, lũng, rặng, sang, sáng; chữ 𠵽 có các

âm sau: lǎi, lǎy, lǎy, lǎy, lǎ, lǎ, lǎ, rǎy, rǎ, rǎ, trǎ, trǎ; chữ 弄 có các âm: lòng, lòng, lộng, lộng, lòng, sổng; chữ 停 có các âm Nôm: dừng, đành, đành, dừng, rành, đình.

Những bước triển khai và tiến hành công việc

Để tiến hành công việc thiết kế cho một dự án tự điển Nôm online, tất cả các yếu tố liên quan đến việc quản lý dự án như nhân lực, phân công, quy trình làm việc, công cụ cần thiết, kỹ thuật, tác phẩm tham khảo, nguyên tắc làm việc, đặc tính cần có, giới hạn đã biết, quy tắc thông tin nội bộ, quy trình làm việc (development process), quy định kỹ thuật (functional specifications), phần Giao diện cho người dùng (Guide User Interface-GUI) và Cơ sở dữ liệu (database) v.v đều phải được viết vào những hồ sơ quản lý dự án (project plans) rõ ràng, phải được ban ban tập cứu xét, sửa chữa nhanh chóng và ban hành cho mọi người hiểu rõ tức thời, nhất là phải bàn luận thận trọng những gì cần phải có trong Quy định kỹ thuật, Cơ sở dữ liệu và phần Giao diện cho người dùng. Về phần Giao diện và Cơ sở dữ liệu thì ít nhất những yếu tố sau đây phải được thỏa mãn: 1) Các phương pháp tra tìm chữ (bằng bộ thủ, bằng nét, bằng âm đọc, bằng chữ Unicode, bằng mã Unicode, bằng tập hợp thành phần chữ), 2) Các dữ kiện cần có cho một chữ Nôm như thông tin về bộ thủ, số mã Unicode, mã nội bộ, số nét, tổng số nét, kết cấu của chữ, các âm đọc và phân loại các ví dụ trích dẫn từ các văn bản, cũng như thông tin về các văn bản đó như tác giả, tên sách, nhà xuất bản, năm xuất bản v.v.

Ban biên tập và thiết kế phải như thế nào ?

Ban Biên Tập phải bao gồm các chuyên gia ngôn ngữ Hán-Việt-Nôm, các nhà nghiên cứu ngữ âm học lịch sử, các chuyên viên kỹ thuật điện toán và Internet, các chuyên gia chế tạo phông chữ và cả những người mới tìm học chữ Nôm. Để công việc tiến hành dễ dàng và có hiệu quả cần phải có những phương cách điều hành thông tin nội bộ làm sao cho thông tin nội bộ được nhanh chóng, hiệu quả và thông suốt. Sự phát triển của kỹ thuật điện toán cùng khả năng truyền thông trên Internet đã cho phép kết hợp một gồm nhiều người cư ngụ tại nhiều nơi trên thế giới, thế nên, còn phải cần đến những công cụ thông dụng để điều hành thông tin nội bộ như Blog, Listservers v.v

Quy trình làm việc.

Công việc thực hiện tự điển online nên được tiến hành và lập đi lập lại nhiều lần như sau:

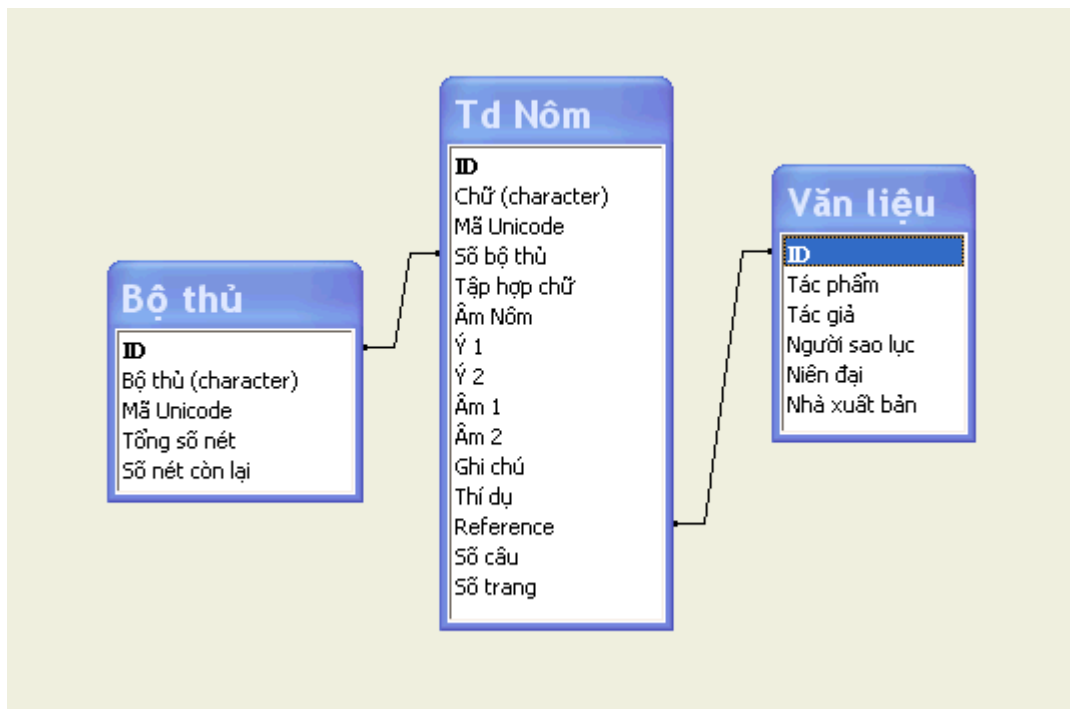
1. Thâu thập các tài liệu, văn bản chữ Nôm từ các thư viện trên thế giới hoặc từ những tư liệu của các nhà nghiên cứu. Chọn lọc một số văn bản nòng cốt để lập thành một "Thư mục" dùng làm cơ sở liệt soát từng chữ Nôm một. Đưa chữ vào tự điển với tinh thần tôn trọng tối đa nguyên tác, các thí dụ được trích dẫn phải chọn lọc kỹ càng và chính xác.
2. Thiết lập Giao diện. Có thể thiết kế một Giao diện riêng cho Ban biên tập có những chức năng căn bản như thống kê, phân tích, tìm kiếm, thay đổi, tra tìm v.v.
3. Thiết lập và cập nhật Cơ sở dữ liệu
Mọi tương quan giữa phần Giao diện và Cơ sở dữ liệu đều phải được kiểm chứng và phải phù hợp với những yêu cầu trong các hồ sơ quản lý dự án.
4. Chế tạo phông chữ Nôm theo đúng tiêu chuẩn mã quốc tế Unicode Standard và Microsoft Specifications for True Type Fonts cho mỗi chữ Nôm của kết quả công trình liệt soát chữ từ các văn liệu sưu tầm được.
5. Thiết kế và cập nhật một bàn gõ để cho công việc gõ nhập chữ được dễ dàng và nhanh chóng hơn.
6. Gõ nhập chữ vào database của tự điển. Mỗi chữ (mục từ) phải được kiểm chứng (testing) nhanh chóng để bảo đảm tính chính xác.

Ngoài quy trình thiết kế một cuốn tự điển online trên đây, còn phải được chú trọng đến sự tương quan giữa quy trình thiết kế bộ phong đang sử dụng và một công cụ bàn gõ phụ giúp thêm cho công việc gõ nhập chữ vào Cơ sở dữ liệu của tự điển Nôm online. Không như các ngôn ngữ châu Á khác như Hán, Hàn và Nhật, các ngôn ngữ này đều có các bộ mã quốc gia đầy đủ, nên việc thiết kế phần mềm cho máy tính ít bị trở ngại vì vấn đề phong chữ. Chữ Nôm thì trái lại. Mặc dù đã có được khoảng 10,000 chữ được Unicode mã hóa, song con số này vẫn chỉ được xem như là nhỏ vì có rất nhiều chữ Nôm chưa có mã Unicode còn nằm nhan nhản trong các văn bản Nôm. Nếu không có một bộ phong đầy đủ thì việc ứng dụng chữ Nôm trên mạng, trên máy tính vẫn còn nhiều hạn chế. Và nếu như không có một công cụ gõ nhập tự điển một cách dễ dàng và hiệu quả thì công việc thiết kế tự điển online cũng sẽ gặp nhiều trở ngại. Do vậy, phải có những liên hệ mật thiết và đồng bộ giữa ba quy trình làm việc này (Cơ sở dữ liệu, Bàn gõ và Phong chữ).

Cấu trúc Cơ sở dữ liệu

Xác định cấu trúc cơ sở dữ liệu ngay từ lúc đầu là ưu tiên hàng đầu đối với các dự án làm tự điển nói chung và tự điển chữ Nôm nói riêng. Việc phân chia các dữ kiện thành từng trường (field) với những quy ước nhất định giúp giảm thiểu các sai sót trong khi nhập dữ kiện và cho phép người dùng tìm kiếm (search), sắp xếp (sort) dữ kiện theo ý muốn một cách nhanh chóng. Ngoài ra, kho dữ liệu còn là nền tảng của các công cụ phân tích, thống kê, nghiên cứu, v.v. Ta còn có thể liên kết kho dữ liệu hiện có với các kho dữ liệu khác để giảm thiểu việc nhập dữ kiện.

Hình vẽ dưới đây là một đề nghị cho một cấu trúc kho dữ liệu căn bản cho Tự Điển Chữ Nôm Trích Dẫn-một dựa án Tự Điển Chữ Nôm online.

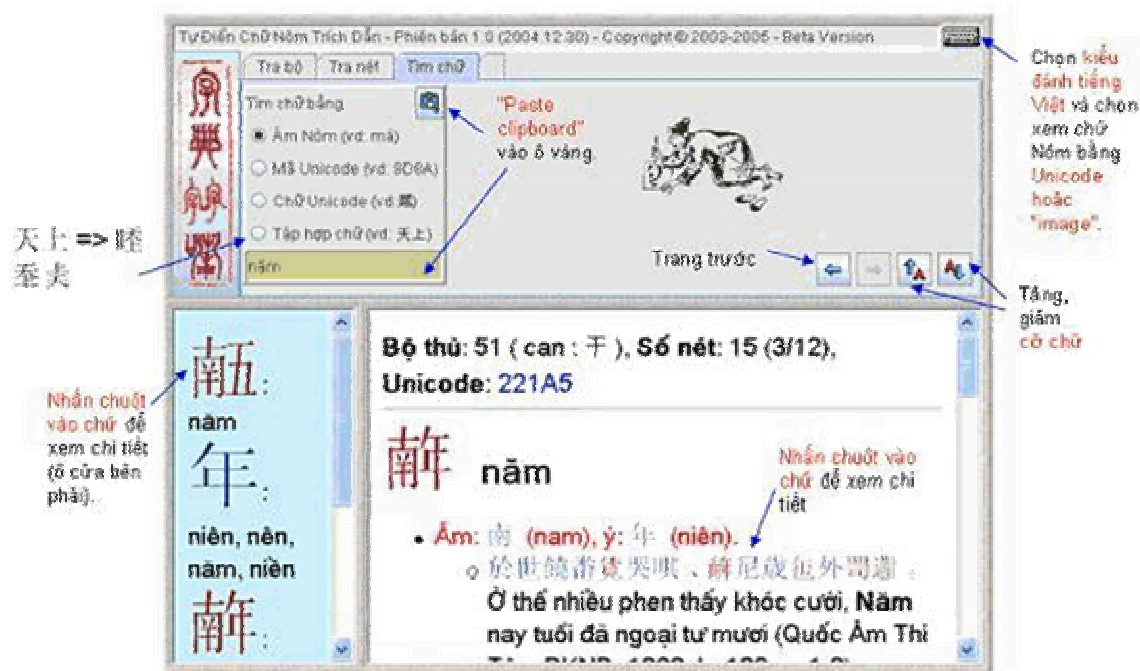


Theo cấu trúc này, người dùng có thể tìm kiếm (search) và sắp xếp các dữ kiện theo bất kỳ trường (field) nào có mặt trong các bảng (table) này (chữ, mã Unicode, tập hợp chữ, âm Hán Việt, âm Nôm, ý phù, âm phù, văn liệu, v.v.). Các bảng (table) có thể được thêm bớt những yếu tố cần thiết để phù hợp với các chức năng thống kê hay tra cứu sau này của tự điển. Ví dụ: Nếu dự định Tự Điển này có thể cho phép người dùng tra tìm chữ Nôm theo phương pháp 4 góc

(Four corner method) thì trong phần cấu trúc TdNôm chúng ta có thể thêm vào các dữ kiện cần thiết của phương pháp 4 góc. Các chức năng thống kê phải được thiết kế theo chiều hướng giúp cho các nhà nghiên cứu có thêm phương tiện tra cứu, cung cấp các dữ kiện quan trọng cho công việc nghiên cứu chữ Nôm, ví dụ: tìm ngữ cảnh của chữ Nôm, hoặc làm thống kê tổng hợp các dạng chữ Nôm có cùng một âm đọc, hoặc so sánh âm nghĩa, phân tích, suy luận hầu xác nhận hoặc đưa ra những kiến giải về cách cấu tạo, diễn biến của sự sử dụng chữ Nôm qua các thời kỳ lịch sử, v.v. Cơ sở dữ liệu của tự điển chữ Nôm online chứa đựng chữ Nôm dựa trên căn bản liệt soát các văn bản, theo một khía cạnh nào đó, chính là một cuốn từ điển chữ Việt cổ.

Giao diện cho người dùng (end-user's interface)

Ngoài việc cung cấp các chức năng cơ bản của một tự điển như tìm kiếm theo âm, mã Unicode, chữ, tập hợp chữ, v.v. ta có thể mở rộng Giao diện cho người dùng bằng cách cung cấp các công cụ (tools) như thống kê, phân tích, phiên âm, bàn gõ, kho văn bản, v.v. mà cơ sở dữ liệu cho phép. Giao diện cần phải gọn gàng, dễ dùng, và tiện lợi. Giao diện cho người dùng (chức năng cơ bản) của Tự Điển Chữ Nôm Trích Dẫn được thiết kế như sau:



Có thể thiết kế thêm một Giao diện dành riêng cho Ban Biên Tập cuốn Tự Điển đó có những chức năng căn bản như thống kê, phân tích, tìm kiếm, thông tin nội bộ v.v. Mọi tương quan giữa phần Giao diện và Cơ sở dữ liệu đều phải được kiểm chứng và phải phù hợp với những yếu tố yêu cầu trong các hồ sơ quản lý dự án.

Vấn đề quản lý chất lượng

Chính xác là điều đòi hỏi hàng đầu mà một Tự Điển cần phải có. Giá trị của một cuốn Tự Điển sẽ bị giảm đi rất nhiều, nếu người dùng phát hiện có nhiều sai lạc, lầm lỗi hiện diện trong cuốn tự điển đó. Sự hiện diện của các sai lầm này sẽ vô cùng tai hại, vì sẽ có nhiều người dùng tự điển sẽ bị sai lầm theo. Do vậy, thiết kế tự điển Nôm online cũng giống như thiết kế các loại phần mềm khác đòi hỏi sự quan tâm và chú trọng đặc biệt đến các quy trình thiết kế, phát triển và kiểm chứng phần mềm, nhất là quy trình kiểm chứng các dữ liệu đưa vào tự điển, không được để

sót 1 lỗi làm nào tồn tại. Áp dụng các phương pháp quản lý chất lượng ứng dụng hiện hành trong kỹ thuật thiết kế và kiểm chứng phần mềm từ khi dự án bắt đầu tiến hành cho tới khi dự án chấm dứt là điều cần thiết và quan trọng. Chất lượng của tự điển online phụ thuộc rất nhiều với chất lượng của các loại hồ sơ quản lý dự án (project plans) như Quy trình làm việc, Quy định kỹ thuật (functional specifications), Giao diện cho người dùng (Guide User Interface-GUI) và Cơ sở dữ liệu (database) v.v. Các hồ sơ quản lý dự án này phải được ban Biên tập bàn luận và kiểm chứng nhiều lần qua suốt thời kỳ tiến hành dự án. Một đội ngũ làm công việc kiểm chứng và phát hiện các lỗi làm làm việc tức khắc ngay sau khi các quá trình thiết kế tiến hành xong. Như vậy, quy trình thiết kế-kiểm chứng (design-test cycles) sẽ xảy ra liên tục cho đến lúc dự án chấm dứt.

Các công cụ cần thiết:

Đề dự án có thể tiến hành nhanh chóng và có hiệu quả cho việc thiết kế một tự điển online, nên thiết kế riêng cho dự án những công cụ cần thiết, nhưng đồng thời nên nghiên cứu và tìm kiếm những công cụ (tools) sẵn có trên thị trường. Những công cụ tra cứu sau đây nên đáng liệt kê vào danh sách cần dùng như:

1. Unicode Character Map Utility for Windows: <http://www.babelstone.co.uk/Software/BabelMap.html>
2. UniHan database: <http://www.unicode.org/charts/unihan.html>
3. A Nôm Lookup Tool based on Unicode <http://nomfoundation.org/unicode/lookups.php>
4. Những phần mềm thông dụng để gõ nhập chữ Hán
5. Bàn gõ chữ Hán/Nôm (HanNomIME) của Alexandre Lê và Trần Uyên Thi
http://www.viethoc.org/hannom/bango_intro.php

Kết Luận

Việc bảo tồn và phát huy những giá trị văn hóa truyền thống tiềm tàng trong di sản chữ Nôm ngày càng được coi trọng và cấp bách. Công nghệ thông tin hiện đại được xem như là một phương cách hữu hiệu để phục hồi và bảo tồn di sản này. Vai trò của một cuốn tự điển chữ Nôm online sử dụng trên mạng Internet cũng theo đó phải có với một diện mạo mới, phong thái mới và kỹ thuật mới để đáp ứng nhu cầu đó. Một tự điển chữ Nôm online phải có một phương pháp tra tìm nhanh chóng và hiệu quả cho phép người dùng có thể tìm kiếm một chữ Nôm bằng thành phần cấu tạo nên chữ Nôm đó mà không cần biết thêm các dữ kiện nào khác; phải có những phương cách giúp đỡ người dùng hiểu được kết cấu tạo nên chữ đó khi dùng ở trong nhiều văn cảnh khác nhau; phải được biên soạn dựa trên cơ sở liệt soát, sao lục và trích dẫn chính xác từ các văn bản Nôm xưa và đồng thời cơ sở dữ liệu (database) của tự điển phải thiết kế như thế nào để có thể cung cấp các dữ kiện dùng để phân tích, thống kê cho các nhà nghiên cứu Nôm; cũng như dễ chuyển đổi (migration) vào các platform khác.

Tài liệu tham khảo

1. Thiều Chửu (1942). *Hán-Việt Tự-Điển* 漢越字典. Hanoi: Đuốc Tuệ
2. Chen Ching-ho 陳荆和 (1949). 喃字之形態與產生之年代. Taipei, Taiwan: 人文科學論叢.
3. Đào Duy Anh (1975). *Chữ Nôm: Nguồn-gốc, cấu-tạo, diễn-biến*. Hà Nội: Khoa Học Xã Hội.
4. 佛說大報父母恩重經, Société Asiatique, 1822
5. 梁實秋. *A new practical Chinese-English dictionary*. 台灣: 遠東圖書.
6. 東京市 (1988). 字喃字典, 竹内与之助著. 大學書林, Tokyo, Japan.
7. Nguyễn Đình Hòa (1992). *Graphemic Borrowings from Chinese - The Case of Chu Nom - Vietnam's Demotic Script*. Taipei, Taiwan: Bulletin of the Institute of History and Philosophy, Volume 61, part 2.

8. Nguyễn Khắc Kham (1974). *Chữ Nôm or the Former Vietnamese Script and Its Past Contributions to Vietnamese Literature*. Tokyo University, Japan: Area and Culture Series, volumn 24.
9. Tự Điển Chữ Nôm Trích Dẫn http://www.viethoc.org/hannom/tdnom_intro.php
10. Đặng Thế Kiệt, Lê Văn Đặng, Nguyễn Hữu Vinh, Tự Điển Hán Việt Thiệu Chử online http://www.viethoc.org/hannom/tdtc_intro.php
11. Trần Uyên Thi và Alexandre Lê, *Bàn gở Hán Nôm (HanNomIME)* http://www.viethoc.org/hannom/bango_intro.php
12. Trần Văn Giáp (2002), “*Lược Khảo Vấn Đề Chữ Nôm*”, Lê Văn Đặng thực hiện văn bản, Ngày Nay Publishing, USA.
13. Nguyễn Quang Hồng (2002), “*Chữ Nôm trên đường hội nhập với khu vực và thế giới*”, Viện Nghiên Cứu Hán Nôm, Hà Nội.
14. Nguyễn Tá Nhí (1998), “*Các Phương Thức Biểu Âm trong cấu trúc Chữ Nôm Việt*”, Nhà Xuất Bản Khoa Học Xã Hội, Hà Nội.
15. Nguyễn Hữu Vinh (2005), “*Chữ Nôm và Tinh thần dân tộc*”, Viện Việt Học, California, USA