

# Nôm Lookup Tool (NLT)

- ≡ Lý do thành hình
- ≡ Đã làm được gì
- ≡ Phải làm thêm gì cho các ấn bản NLT trong thời gian tới
- ≡ Các hướng nghiên cứu khác bổ túc vào NLT
- ≡ Được gì khi tham gia nhóm
- ≡ Kết luận

# Lý do thành hình (1)

- ⇒ Prototype để hiểu các yêu cầu về giao diện người dùng, cách tìm nào là dễ dàng, chính xác và nhanh nhất
- ⇒ Công cụ giúp đỡ việc mã hoá chữ Nôm
  - ⇒ có những thông tin chuyên môn không nhất thiết người sử dụng cuối cần. Thí dụ: mã Unicode, mã tài liệu dùng
- ⇒ Mô hình dữ liệu để hiểu dần dần cụ thể về cấu trúc một từ điển số chữ nôm

## Lý do thành hình (2)

- ⇒ "Vật" cụ thể để đối thoại giữa kỹ thuật viên không có kiến thức chữ Nôm và các nhà chuyên môn trong nhóm
- ⇒ Công cụ để kỹ thuật viên hiểu biết thêm cấu trúc, đặc thù tiếng Nôm
- ⇒ Mục đích cuối của nhóm là tạo **điều kiện dễ dàng nhất cho việc học chữ Nôm**, thí dụ: cách tìm chữ Nôm thích hợp
- ⇒ Nhóm chưa có ý định làm ra sản phẩm thương mại.

## Lý do thành hình (3)

- ⇒ Nhân sự: một nhóm bạn làm nhiều chuyện chung trong đó có hai lĩnh vực chủ yếu: đào tạo về CNTT, chuẩn tiếng Việt (hiện nay tập trung về chữ Nôm)

# Đã làm được gì – NLT 2001 (1)

## ☰ Yêu cầu:

- ☑ Phục vụ những người không biết, hoặc biết rất ít chữ nôm
- ☑ Chưa biết giải quyết ra sao về mặt kỹ thuật của việc nhập chữ nôm bằng một bàn phím latin

## ☰ Quyết định:

- ☑ Đầu vào là các từ quốc ngữ

## ☰ Nhận xét:

- ☑ Một từ quốc ngữ tương ứng nhiều từ nôm. Càng có nhiều chi tiết, kết quả tìm càng chính xác

# NLT 2001 (2)

## ☰ Quyết định:

- ☑ có thể tìm theo âm (sound), biểu ý (radical), âm và biểu ý, nghĩa (hiện thời tiếng Anh vì nó có sẵn khi nộp cho Unicode), mã Unicode

## ☰ Ghi chú:

- ☑ trong bài qui trình Nôm Na dùng khái niệm đơn vị chính tả thay cho khái niệm "radical, sound" ở NLT

## ☰ Kết quả:

- ☑ Prototype đã trình bày tại Hội Thảo Hè Aix 2001

# NLT 2001- Giới hạn (1)

- ☐ Người dùng phải có phong chữ tiếng Nôm.
  - ☐ Do vấn đề giấy phép và bản quyền, phong chữ Nôm thử nghiệm chỉ được quyền dùng cho mục đích nghiên cứu
  - ☐ Thành viên của nhóm phải thực hiện nghiêm chỉnh qui định không được phép phổ biến tự do phong chữ Nôm
  - ☐ Việc này làm giới hạn phổ biến NLT 2001. Một số giới hạn người ngoài nhóm được biết và thử NLT như giáo sư Nguyễn Quang Hồng

## NLT 2001- Giới hạn (2)

- ☐ Phong thử nghiệm hiện nay không tuân thủ đúng mã Unicode
- ☐ Do chưa đủ dữ liệu nên người dùng chưa tìm được nhiều
  - ☐ Biết được "trời" là gồm "thiên" và "thượng" nhưng không biết "thiên" và "thượng" viết ra sao nên không nhận ra được ghép ra sao
  - ☐ Không có thông tin về số nét
  - ☐ Thiếu thông tin về gốc của từ : Hán-Việt hay thuần Việt

## NLT 2001- Giới hạn (3)

- ☐ gái = "nữ" + "cái" nhưng lại không tìm được chữ "nữ" để biết viết ra sao ?
- ☐ Chưa có ý kiến nhiều của người dùng do giới hạn việc phổ biến thông
- ☐ Một số khuyết điểm cơ bản chỉ khắc phục được khi dữ liệu trữ đầy đủ hơn

# NLT 2002-1: Vấn đề

## ☰ Vấn đề:

- ☑ Làm cách nào hiển thị chữ nôm mà không cần có phong chữ nôm
- ☑ Phong nôm làm ra tốn kém nhưng làm xong có ai mua không ? Ngoài một số người có chủ đích nghiên cứu về văn hoá, nôm học, những người có nhu cầu rất đơn giản (muốn biết tên mình viết ra sao, chữ nôm có chữ tục không, tên sách, tên phim, vài câu chúc, ...) liệu họ có mua phong nôm không ?

# NLT 2002-1: Lời giải (1)

## ☰ Lời giải:

- ☑ Anh [James] Đỗ Bá Phước đã áp dụng kỹ thuật Scalable Vector Graphics, một chuẩn mới của W3C
- ☑ Chuẩn này nhằm áp dụng để hiển thị những dữ liệu ở dạng sơ đồ graphic như trong thống kê xác suất (hình bánh, cột, ...). Áp dụng chuẩn trên, 9299 ký tự nôm được chuyển thành dữ liệu SVG.  
Hiện nay để hiển thị các dữ liệu SVG với các duyệt trình IE 6, Netscape 6.2 cần cài thêm một plugin miễn phí.
- ☑ SVGViewer của Adobe (có thể tưởng tượng nó như Adobe Reader)
- ☑ Như thế nhu cầu đơn giản nói trên đã được giải quyết. Sau khi phổ biến URL số người vào trạm của nôm foundation tìm chữ Nôm tăng vọt

## NLT 2002-1: Lời giải (2)

- Trong tương lai gần các duyệt trình sẽ hỗ trợ SVG và SVG Viewer sẽ là một thành tố của duyệt trình. Nếu SVG cũng được hỗ trợ bởi các phần mềm Office thì điều kiện hiển thị chữ Nôm sẽ rất thuận tiện. (Đã đúng cho Adobe SVG Viewer)

## NLT 2002-2: bổ túc nhỏ

- ⇒ hiển thị unicode khi gõ các chữ quốc ngữ: chạy tốt cho IE nhưng không có kết quả cho NS
- ⇒ tạm giữ cách gõ (Input Method): VIQR

## Các ấn bản mới (1)

- ⇒ Thêm dữ liệu: trên kết quả của dự án qui trình Nôm Na, NLT sẽ được bổ túc nhiều thông tin hơn việc tìm chữ nôm thuận tiện hơn
- ⇒ Với một bộ phong chữ Nôm ra đời từ dự án Nôm Na cho phép việc phổ biến NLT dễ dàng hơn

## Các ấn bản mới (2)

- ⇒ Tùy thuộc vào kết quả nghiên cứu của các nhà ngôn ngữ học về cách tìm và yêu cầu của người sử dụng, NLT sẽ bổ túc thêm những cách tìm chữ khác
- ⇒ Có giao diện người dùng để nhập dữ liệu

## Các hướng nghiên cứu khác

- ⇒ Tìm từ quốc ngữ qua chữ nôm.  
Điều kiện: có cách gõ tiếng Nôm qua bàn phím ký tự la tinh
- ⇒ Sử dụng kỹ thuật SVG để tạo chữ mới

# Được gì khi tham gia nhóm

## ☰ Vui

- ☰ Thiệp chúc Tết
- ☰ Tìm chữ theo ý thích: Em chết mất, như cánh vịt quay

## ☰ Hiểu thêm cấu trúc chữ nôm

## ☰ Quen biết thêm những người am hiểu chữ nôm, ngôn ngữ học

## ☰ Có dịp thảo luận những kỹ thuật mới kể cả trong tin học

# Kết luận

- ⇒ NLT là một việc chưa làm xong
- ⇒ Mỗi chặng đường có thêm những câu hỏi mới

# Thí dụ (1)

⇒ <http://nomfoundation.org/unicode/lookup.php>

⇒ Em chết mát

⇒ 俺 斃 忙

⇒ Em 俺 nữ yêm

Chết 斃 tử 死 chiết 折

Mát 忙 vong 亡 mạt 末

## Thí dụ (2)

- ⇒ như cánh vịt ngỗng chim quay bay  
如 翹 鶻 鷓 鴒 插 飛
- ⇒ như: 如  
cánh 翹: vũ 雨 canh 更  
vịt 鶻: u+7225??? việt 越
- ⇒ điều 鳥, việt 越 (GDNHV 935.11)
- ⇒ ngỗng 鷓: 鳥 điều 貢 nga 鵝?  
chim 鷓: điều 鳥 chiêm 占  
quay 插: thủ 手 quai 乖  
bay 飛: phi 飛 bi 悲