

# **Textual issues of 大越史記全書 *Đại Việt Sử ký toàn thư* [A Complete History of the Dai Viet]**

Ngô Thanh Nhân (*Temple University*) & Ngô Trung Việt (*Institute of Informatics*)  
Lê Văn Cường, Tô Trọng Đức, Lương Thị Hạnh (*Nôm Na Group, Hanoi*)

## **ABSTRACT**

Creation of an electronic version of 大越史記全書 *Đại Việt Sử ký toàn thư* (ĐVSKTT, A Complete History of the Great Viet) is no longer a problem thanks to the inclusion Nôm in the multi-lingual computer standards. However, accessing the digitized ĐVSKTT raises several basic textual questions: such as dates, names, basic syntactic structures and the general principles of historical events stated in its 纂修九例! *Soán tu phàm lệ* [compilation convention]. Dates must be identified and converted to and from the Gregorian system [standard W3C Date Time Format]. Names include geographic names, plant, animal and object names, titles, and personal names of various styles. Geographic names can be administrative (assigned by different regimes and their borders), local and original names. Personal names include names at birth, pen names, nicknames, titles, diminutives, taboos, ... A name usually has many forms cited above including “pronouns”, which can be construed as context-sensitive and time-sensitive “synonyms”. For example, the title 太宗 *Thái tông* refers to different king at different time in Vietnamese history. Perhaps, taboo lists may be consulted as a matter of protocol.

This paper presents some basic textual problems of ĐVSKTT in web access, drawn from the most common textual description used by Nôm scholars. They may be engineered as discovery procedures for properties of Nôm texts. The first application is aimed at library information search routines using controlled vocabulary and textual analyses, such as ideogram count, word count, sentence/predicate identification, border words,...

## Những vấn đề văn bản học của 大越史記全書 *Đại Việt Sử ký toàn thư*

Ngô Thanh Nhân (*Đại học Temple*) & Ngô Trung Việt (*Viện Công nghệ Thông tin*)  
Lê Văn Cường, Tô Trọng Đức, Lương Thị Hạnh (*Nhóm Nôm Na*)

### Tóm tắt

Số hoá nguyên văn 大越史記全書 *Đại Việt Sử ký toàn thư* (ĐVSKTT) ngày nay không còn là một vấn đề nữa nhờ việc quốc tế hoá mã chữ Nôm. Để sử dụng văn bản số hoá như ĐVSKTT, một số những vấn đề văn bản cần đặt ra để giúp người sử dụng vượt qua các trở ngại truy cập: đó là vấn đề ngày tháng, tên, ngữ pháp cơ bản và những vấn đề nguyên tắc ghi trong *Phàm lệ về việc biên soạn ĐVSKTT*. Ngày tháng cần được nhận diện và chuyển đổi từ hệ thống âm lịch Việt Nam sang hệ dương lịch và ngược lại. Tên gồm có địa danh, cây cỏ, niên hiệu, tên họ nhân vật. Tên địa danh gồm tên địa danh hành chính (tên của các vì vua đặt ra và biên giới của chúng), tên địa phương và tên nguyên thủy. Tên người gồm có tên thật, tên hiệu, bút hiệu, chức tước, tên gọi trong gia đình, tên húy, v.v. Một người có nhiều tên và nhiều cách gọi cũng như đại từ quan hệ gia đình. Có lẽ, khi tìm một tên, một cách trả lời gồm những tên khác nhau (của cùng một người hay nhiều người) cùng hiện ra theo thời điểm xuất hiện của chúng, ví dụ tìm 太宗 *Thái tông*. Và vì hai lẽ trên, có lẽ cần có *các bảng húy* và thời điểm chúng tác động vào văn bản.

Bài này giới thiệu sơ khởi quy trình văn bản Nôm học trong công tác số hoá ĐVSKTT để tra cứu mạng. Quy trình này dựa trên những “thao tác” thông dụng nhất của văn bản học, và có thể triển khai thành quy trình cơ sở cho các văn bản số hoá cho mọi loại ứng dụng trên mạng. Ứng dụng đầu tiên là ngành thư viện (tra cứu theo các bảng từ ngữ định sẵn) và nghiên cứu văn bản.