

# GIẢI PHÁP TIN HỌC HÓA TỪ ĐIỂN THỐNG KÊ VIỆT NAM PHỤC VỤ TRA CỨU VÀ CẬP NHẬT TRỰC TUYẾN

ThS. Vũ Thị Vân Anh\*

## **Tóm tắt:**

*Từ điển Thống kê được Tổng cục Thống kê triển khai biên soạn từ năm 2012 và xuất bản tháng 1/2016, kế thừa cuốn Từ điển xuất bản năm 1977. Với mong muốn cuốn Từ điển Thống kê được phổ biến rộng rãi, thuận tiện trong việc tra cứu và cập nhật nội dung, cần xây dựng cuốn từ điển điện tử, bài viết này đề cập đến các giải pháp tin học hóa cuốn Từ điển Thống kê.*

Từ điển Thống kê gồm 2300 từ, thuật ngữ. Mỗi thuật ngữ được trình bày ngắn gọn, đầy đủ, rõ ràng khái niệm và những nội dung cơ bản. So với các từ điển chuyên ngành đây là cuốn từ điển chứa đựng nội dung kiến thức sâu và rộng đối với chuyên ngành thống kê và các ngành liên quan. Nhằm tạo điều kiện thuận tiện cho việc cập nhật nội dung từ điển và làm điều kiện cho việc tiếp cận Từ điển Thống kê, cần thiết phải tin học hóa cuốn Từ điển Thống kê, chúng tôi đề xuất 3 giải pháp sau:

### **Thứ nhất, giải pháp về công nghệ**

Trên cơ sở đánh giá hiện trạng công nghệ thông tin của Tổng cục Thống kê và Viện Khoa học Thống kê, chúng tôi đề xuất công nghệ và thiết bị đi kèm phải đồng bộ. Vì để đảm bảo số lượng lớn người dùng truy cập mà không bị nghẽn máy chủ và hệ thống xử lý, các dịch vụ trên web bao gồm các ứng dụng xử lý trực tuyến trên web, các ứng dụng truy cập, chỉnh sửa, biên tập trực tuyến ngày càng gia tăng để giải quyết vấn đề đó thì một hệ thống quản lý và phân quyền người sử dụng phải được xây dựng và quản lý một cách thống nhất.

- Công nghệ quản trị hệ thống thông tin, gồm: An toàn, bảo mật thông tin, cơ sở dữ liệu, bảo mật thông tin truyền trên mạng, khuyến nghị sử dụng các phương pháp mã hoá, cập nhật dữ liệu. Vì vậy, một số giải pháp:

+ Lựa chọn kiến trúc công nghệ thông tin phục vụ chương trình chạy Phần mềm tra cứu Từ điển Thống kê trực tuyến;

+ Hoàn thiện cơ sở hạ tầng công nghệ thông tin: Hiện tại, máy chủ cần phải nâng cấp nhằm tăng khả năng lưu trữ dữ liệu cho hệ thống; xây dựng thêm hệ thống dự phòng; toàn bộ hệ thống chưa được bảo mật firewall nên cần xây dựng ngay; toàn bộ hệ thống chưa được trang bị thiết bị lưu trữ dự phòng để phòng các sự cố và rủi ro; hệ thống đường dẫn là rất quan trọng đối với mô hình lưu trữ dữ liệu tập trung;

+ Lựa chọn công nghệ, phần mềm để xây dựng Web Từ điển Thống kê trực tuyến: Hệ quản trị dữ liệu SQL Server. Hiện nay Tổng cục Thống kê đang sử dụng hệ quản trị SQL Server và định hướng giai đoạn 2016-2020 vẫn tiếp tục sử dụng hệ quản trị này. Nếu Web tra cứu Từ điển Thống kê đặt

\* Phó Trưởng phòng Nghiên cứu khoa học và Chiến lược phát triển Thống kê, Viện Khoa học Thống kê

tại web

<http://www.gso.gov.vn/Default.aspx?tabid=217>  
thì lựa chọn SQL Server là ưu tiên hàng đầu.

Tuy nhiên, dự kiến Web tra cứu do Tổng cục Thống kê quản lý thì thuận lợi về hạ tầng công nghệ nhưng lại bất cập trong việc quản lý vận hành, người theo dõi và cập nhật khi mà Viện Khoa học Thống kê từ khi triển khai xây dựng Từ điển đã được giao là đơn vị chủ trì. Nếu Web tra cứu tại Viện Khoa học Thống kê, thì đảm bảo hạ tầng công nghệ thông tin cần được cải thiện nhiều.

**Thứ hai, giải pháp nghiệp vụ**

Một trong những nội dung quan trọng nhất để tin học hóa Từ điển Thống kê là nội dung của Web tra cứu. Đưa những gì của Từ điển Thống kê vào để tra cứu là một câu hỏi định hướng rà soát Từ điển Thống kê? Rà soát Từ điển Thống kê theo những trường

nào để đáp ứng web tra cứu không bị lạc hậu, nội dung và hình thức phong phú.

Trên cơ sở nghiên cứu kinh nghiệm của một số web tra cứu Từ điển nổi bật và đề xuất rà soát Từ điển Thống kê theo những nội dung: (i) Số lượng từ, thuật ngữ Từ điển Thống kê phân theo vần Anphabeta; (ii) Số lượng từ, thuật ngữ Từ điển Thống kê phân theo cấu trúc từ, thuật ngữ; (iii) Từ, thuật ngữ Từ điển Thống kê gần giống tên các chỉ tiêu thống kê; (iv) Từ, thuật ngữ Từ điển thống kê phân theo nguồn tài liệu. Tất cả những trường dữ liệu này phục vụ mục đích xây dựng cơ sở dữ liệu đáp ứng tính năng của Web tra cứu. Mặt khác, kết quả của mỗi lần tra cứu là từ, thuật ngữ với các nội dung cơ bản hiển thị về từ, thuật ngữ (gọi chung là dữ liệu về nội dung của từ, thuật ngữ hiển thị). Dự kiến khi tra cứu ta thu được kết quả như Bảng 1.

**Bảng 1:** Nội dung dự kiến hiển thị trên Web tra cứu cuốn Từ điển Thống kê

Tên từ, thuật ngữ	Tên tiếng anh	Nội dung của từ, thuật ngữ trong Từ điển Thống kê	Lĩnh vực liên quan	Bối cảnh của định nghĩa	Nguồn tham khảo	Các link liên kết với những định nghĩa có liên quan
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Từ, thuật ngữ 1						
...						
Từ, thuật ngữ 2300						

Không chỉ có vậy, những dữ liệu khác cần được biên tập để phục vụ tính năng Web tra cứu là:

- Dữ liệu về danh mục lĩnh vực trong từ điển, phục vụ phần hiển thị trên giao diện tra cứu;

- Dữ liệu về danh mục từ, thuật ngữ theo Bảng chữ cái;

- Dữ liệu về giới thiệu, hướng dẫn tra cứu, dữ liệu danh mục người dùng;

- Dữ liệu phục vụ tìm kiếm nâng cao;

- Dữ liệu về quản trị hệ thống.

Như vậy, để có nội dung từng cột, đòi hỏi biên soạn công phu và sáng tạo.

+ Ngoài ra, để phục vụ việc cập nhật, cần chú ý hoàn thiện một số nội dung sau mà đề tài đã sơ bộ đưa ra:

- Rà soát các lỗi chính tả, cập nhật nội dung Từ điển;

- Rà soát tên tiếng anh các từ, thuật ngữ;
- Bổ sung thêm từ thuật ngữ thống kê;
- + Cũng trên cơ sở đó, các tính năng dự kiến của Web tra cứu gồm: (1) Tìm kiếm theo từ, cụm từ và Tìm kiếm nâng cao; (2) Hiển thị từ, thuật ngữ theo Bảng chữ cái; (3) Hiển thị từ, thuật ngữ theo Lĩnh vực thống kê; (4) Truy cập; (5) Đánh dấu, lưu lại, xem lại; (6) Hỏi đáp; (7) Quản trị dữ liệu; (8) Liên hệ; (9) Một số tính năng khác, như: Tính năng cho phép người dùng từ tạo danh mục và danh sách từ của riêng mình; tự động tải các dữ liệu về máy để người dùng có thể tra cứu và sử dụng ngay cả khi không có kết nối mạng; tính năng lưu dữ liệu liên quan Từ điển Thống kê (phần này quan trọng, lưu giữ tài liệu liên quan từ lúc bắt đầu xây dựng Từ điển Thống kê đến lúc phát hành).

**Hình 1:** Giao diện dự kiến phần mềm tra cứu trực tuyến Từ điển Thống kê



**Thứ ba, giải pháp tổ chức thực hiện**

(1) Thành lập nhóm Từ điển Thống kê, thuộc Viện Khoa học Thống kê

(2) Tư liệu hóa tài liệu trong quá trình xây dựng Từ điển Thống kê

Trong 4 năm từ 2012 đến 2015 triển khai biên soạn cuốn Từ điển Thống kê có nhiều phiên bản cuốn từ điển đã được biên soạn, nhiều ý kiến góp ý của chuyên gia về cuốn từ điển, nhiều lần hội nghị, hội thảo về tổ chức thực hiện biên soạn Từ điển Thống

kê. Tất cả đều cần được tư liệu hóa và lưu giữ lại làm tài liệu cho lần tái bản lần sau.

(3) Xây dựng quy chế người truy cập

(4) Đảm bảo đủ nguồn lực cho hoạt động phần mềm tra cứu

Để thuận lợi nhất cho việc quản lý và vận hành Phần mềm, cần lựa chọn một đơn vị thuộc TCTK để quản lý. Đơn vị quản lý có vai trò như admin, can thiệp vào các tính năng có của phần mềm. Đặc biệt, là đơn vị chủ trì cập nhật Từ điển Thống kê khi có chỉ đạo. Đối tượng sử dụng của Phần mềm, trước tiên hướng đến các cán bộ ngành Thống kê, việc truy cập phần mềm bằng tài khoản email nội bộ của ngành. Ngoài ra, để mở rộng đối tượng sử dụng và quảng bá ngành Thống kê, nên hướng đến những người sử dụng ngoài ngành. Tuy nhiên, để đảm bảo tính bảo lập cần thiết lập hệ thống truy cập. Duy trì, vận hành Từ điển Thống kê bao gồm 2 yếu tố: (1) Bảo trì, nâng cấp phần mềm theo định kỳ; (2) Duy trì, nâng cấp nội dung hay cơ sở dữ liệu và các tính năng của Phần mềm.

(5) Nguồn kinh phí

Nguồn kinh phí xây dựng và thực hiện triển khai xây dựng phần mềm tra cứu do Ngân sách nhà nước bảo đảm, được bố trí theo phân cấp quản lý ngân sách nhà nước hiện hành.

(6) Hợp tác với các đơn vị về vận hành, quản lý các phần mềm thống kê

Tổng cục Thống kê có vận hành các phần mềm tra cứu như: Phần mềm danh mục hành chính; danh mục nghề nghiệp; phần mềm tra cứu Từ điển 3 thứ tiếng. Cần có mối liên hệ hợp tác học hỏi kinh nghiệm, phương thức triển khai để phù hợp với quy định chung của Ngành.

**(Xem tiếp trang 14)**

-----

## ***Tiếp theo trang 17***

### ***Tài liệu tham khảo:***

(1) CN. Hoàng Minh Thiện (2014), "*Củng cố kho thông tin tư liệu khoa học thống kê kết hợp tin học hóa*";

(2) PGS.TS. Hoàng Đức Liên và Nguyễn Hữu Ty (2011), "*Giải pháp xây dựng các bộ sưu tập tài liệu số*";

(3) TS. Thiều Văn Tiến (2005), "*Nghiên cứu và đề xuất giải pháp công nghệ*

*tin học hóa công tác xử lý số liệu thống kê tháng, quý, năm do các Cục Thống kê thực hiện*";

(4) TS. Nguyễn Thị Bạch Tuyết (2011) "*Tin học hóa quá trình quản lý hồ sơ tư vấn cho các công ty tư vấn Bộ Xây dựng*";

(5) Tổng cục Thống kê (2016), "*Từ điển Thống kê*";

(6) <http://stats.oecd.org/glossary/>