

ĐẶC ĐIỂM CẤU TRÚC DỮ LIỆU CỦA CƠ SỞ DỮ LIỆU QUỐC GIA VỀ KHOA HỌC VÀ CÔNG NGHỆ VÀ YÊU CẦU ĐỐI VỚI PHẦN MỀM QUẢN TRỊ^(*)

ThS Nguyễn Hồng Hạnh
Cục Thông tin KH&CN quốc gia

Tóm tắt: Giới thiệu một số đặc điểm cấu trúc dữ liệu của cơ sở dữ liệu quốc gia về khoa học và công nghệ. Đề xuất yêu cầu đối với phần mềm quản trị cơ sở dữ liệu quốc gia về khoa học và công nghệ trên cơ sở các đặc điểm cấu trúc dữ liệu nói trên.

Từ khóa: CSDL quốc gia; KH&CN; phần mềm máy tính; chương trình máy tính.

Characteristics of the data structures in the national database on S&T and requirements for the database management software

Abstract: The article introduces characteristics of the data structures in the national database on S&T. It highlights the requirements for the database management software based on these characteristics.

Keywords: National database; science and technology; computer software; computer program.

1. Đặt vấn đề

Cơ sở dữ liệu (CSDL) quốc gia về khoa học và công nghệ (KH&CN) là hệ thống thông tin quốc gia về tiềm lực KH&CN, tích hợp các nội dung thông tin chủ yếu về tiềm lực KH&CN của đất nước, gồm: thông tin về các tổ chức KH&CN; thông tin về nhân lực nghiên cứu khoa học và phát triển công nghệ (NC&PT); thông tin về các nhiệm vụ KH&CN và tình hình ứng dụng kết quả thực hiện các nhiệm vụ KH&CN; thông tin về các công bố khoa học và trích dẫn khoa học; thông tin thống kê KH&CN; thông tin về công nghệ, công nghệ cao và chuyển giao công nghệ; thông tin về KH&CN trong khu vực và trên thế giới; thông tin về sở hữu trí tuệ (SHTT); thông tin về tiêu chuẩn-đo lường-chất lượng (TC-ĐL-CL); thông tin về doanh nghiệp

KH&CN [1]. Với một khối lượng thông tin khổng lồ như trên, việc xây dựng và quản lý CSDL đòi hỏi phải giải quyết rất nhiều vấn đề thiết yếu, trong đó có những vấn đề liên quan mật thiết với nhau. Bài viết đề cập đến vấn đề có thể được coi là một trong những yếu tố cốt lõi quyết định sự tồn tại và phát triển của CSDL quốc gia về KH&CN. Đó là phần mềm quản trị CSDL. Tuy nhiên, do nội dung yêu cầu đối với phần mềm quản trị CSDL quốc gia về KH&CN rất rộng so với giới hạn khuôn khổ bài viết nên chúng tôi chỉ nêu một số yêu cầu cơ bản đối với phần mềm này trên cơ sở đặc điểm cấu trúc dữ liệu của CSDL quốc gia về KH&CN.

2. Đặc điểm cấu trúc dữ liệu cơ sở dữ liệu quốc gia về khoa học và công nghệ

CSDL quốc gia về KH&CN là một hệ thống

(*) Bài viết này là một phần kết quả nghiên cứu của đề tài nghiên cứu KH&CN cấp bộ "Nghiên cứu đề xuất cấu trúc, tiêu chí cho CSDL quốc gia về KH&CN" do Cục Thông tin KH&CN quốc gia chủ trì thực hiện.

thông tin tích hợp về KH&CN, được tạo lập từ các CSDL thành phần với các đối tượng bao quát và nội dung thông tin khác nhau. Đây là một trong những yếu tố cơ bản tạo nên đặc điểm cấu trúc dữ liệu của CSDL. Dưới đây là những đặc điểm chủ yếu trong cấu trúc dữ liệu của CSDL quốc gia về KH&CN.

a) Khối lượng trường dữ liệu lớn

Với những nội dung thông tin như đã trình bày, dự kiến CSDL quốc gia về KH&CN bao gồm 10 CSDL thành phần với khối lượng trường

dữ liệu được đề xuất như trong Bảng 1.

Đặc điểm nói trên nếu không được chú trọng khi xây dựng phần mềm quản trị CSDL, sẽ phát sinh những trở ngại khi xây dựng và vận hành CSDL, có thể sẽ không đáp ứng được các yêu cầu cơ bản về vận hành, khai thác, bảo đảm an toàn an ninh thông tin cho CSDL quốc gia về KH&CN.

b) Sự trùng lặp của một số trường dữ liệu

Có hai trường hợp trùng lặp một số trường dữ liệu trong CSDL quốc gia về KH&CN:

Bảng 1. Số lượng trường dữ liệu các CSDL thành phần thuộc CSDL quốc gia về KH&CN

TT	Tên CSDL thành phần	Quan hệ cấu trúc với CSDL quốc gia về KH&CN	Số lượng trường dữ liệu ⁽¹⁾	Ghi chú
1	CSDL nhiệm vụ KH&CN	Tích hợp vào CSDL quốc gia về KH&CN	33 trường	Cục TTKHCNQG xây dựng và quản lý
2	CSDL tổ chức KH&CN	-nt-	33 trường	-nt-
3	CSDL nhân lực NC&PT	-nt-	28 trường	-nt-
4	CSDL công bố khoa học và trích dẫn khoa học (CSDL tài liệu KH&CN)	-nt-	31 trường	-nt-
5	CSDL thống kê KH&CN	-nt-	33 trường	-nt-
6	CSDL thông tin về KH&CN trong khu vực và trên thế giới (CSDL KH&CN quốc tế)	-nt-	105 trường cho 04 CSDL thành phần ⁽²⁾	-nt-
7	CSDL công nghệ, công nghệ cao và chuyển giao công nghệ (CSDL thông tin công nghệ)	-nt-	Chưa xác định	-nt-
8	CSDL về sở hữu trí tuệ (CSDL SHTT)	Liên kết với CSDL quốc gia về KH&CN	Do Cục SHTT quy định	Cục SHTT xây dựng và quản lý
9	CSDL về tiêu chuẩn-đo lường-chất lượng (CSDL TC-ĐL-CL)	-nt-	Do Tổng cục TC-ĐL-CL quy định	Tổng cục TC-ĐL-CL xây dựng và quản lý
10	CSDL về doanh nghiệp KH&CN (CSDL doanh nghiệp KH&CN)	-nt-	Do Cục Phát triển thị trường và doanh nghiệp KH&CN quy định	Cục Phát triển thị trường và doanh nghiệp KH&CN xây dựng và quản lý

(1) Thông tin về số lượng trường được trích rút từ kết quả nghiên cứu của đề tài cấp bộ ""Nghiên cứu đề xuất cấu trúc, tiêu chí cho CSDL quốc gia về KH&CN" do Cục Thông tin KH&CN quốc gia chủ trì [4].

(2) Số lượng trường được đề xuất cho từng CSDL thành phần của CSDL KH&CN quốc tế xin mời xem trong báo cáo tổng hợp kết quả đề tài nghiên cứu KH&CN cấp nhà nước, mã số KX06.02 do Cục Thông tin KH&CN quốc gia chủ trì [3]

+ *Trùng lặp cả tên gọi và nội dung dữ liệu của trường dữ liệu trong phạm vi CSDL thành phần:*

Trường hợp này xảy ra khi trong cùng một CSDL thành phần có những trường dữ liệu trùng lặp cả tên gọi và nội dung thông tin do dữ liệu trong CSDL được tổ chức theo các mô-đun phù hợp với từng giai đoạn phát triển của đối tượng mà CSDL bao quát. Tiêu biểu cho trường hợp này là CSDL nhiệm vụ KH&CN. Theo quy định, dữ liệu về nhiệm vụ KH&CN được chia thành ba khối tương ứng với ba trạng thái của nhiệm vụ KH&CN là: nhiệm vụ KH&CN đang tiến hành; kết quả thực hiện nhiệm vụ KH&CN và ứng dụng kết quả vào thực tiễn [4]. Như vậy, trong CSDL nhiệm vụ KH&CN có thể có ba mô-đun tương ứng với ba trạng thái nhiệm vụ KH&CN nói trên. Ba mô-đun này sẽ có một số trường dữ liệu trùng nhau cả về tên gọi và nội dung thông tin. Ví dụ:

- Tên nhiệm vụ;
- Cơ quan chủ trì nhiệm vụ;
- Cá nhân chủ trì nhiệm vụ (chủ nhiệm đề tài);
- v.v...

Điều cần lưu ý là khi hiển thị thông tin tìm kiếm, các trường trùng lặp trên cũng không thể thiếu đối với bất kỳ mô-đun nào.

+ *Trùng lặp nội dung thông tin của trường dữ liệu giữa các CSDL thành phần:*

Trường hợp này xảy ra khi các CSDL thành phần có những trường dữ liệu tuy tên gọi khác nhau nhưng có thể có nội dung thông tin giống nhau. Tiêu biểu cho trường hợp này là các CSDL: nhiệm vụ KH&CN, tổ chức KH&CN và nhân lực NC&PT. Ví dụ: trường dữ liệu “Cơ quan chủ trì nhiệm vụ KH&CN” của CSDL nhiệm vụ KH&CN có nội hàm giống trường dữ liệu “Tên tổ chức KH&CN” của CSDL tổ chức KH&CN hoặc trường dữ liệu “Tên tổ chức nơi nhân lực đang làm việc” của CSDL nhân lực NC&PT”; v.v...

Đặc điểm nói trên nếu được chú trọng và giải quyết tốt khi xây dựng phần mềm, sẽ tiết kiệm được thời gian và công sức cập nhật, chỉnh sửa, bổ sung dữ liệu trên cơ sở kế thừa dữ liệu có

sẵn của các trường dữ liệu trùng lặp.

c) *Mối liên kết mật thiết về thông tin giữa một số trường dữ liệu*

Một trong những đặc điểm nổi bật trong cấu trúc dữ liệu của CSDL quốc gia về KH&CN là mối liên kết mật thiết về thông tin giữa một số trường dữ liệu của các CSDL thành phần. Mối liên kết này được thực hiện thông qua các đường kết nối (đường link) từ một trường dữ liệu nào đó của một CSDL thành phần đến một hoặc nhiều trường dữ liệu của một hoặc nhiều CSDL thành phần khác. Các trường dữ liệu được kết nối là những trường có chức năng thông tin tương tự và/hoặc liên quan đến trường dữ liệu đầu tiên. Mối liên kết giữa các trường dữ liệu nói trên tạo điều kiện cho việc cập nhật CSDL cũng như mở rộng phạm vi tìm kiếm để có được thông tin một cách hệ thống. Mối liên kết dữ liệu giữa các trường dữ liệu CSDL quốc gia về KH&CN được thực hiện theo hai phương thức như sau:

- *Liên kết trong cập nhật, chỉnh sửa dữ liệu:* dữ liệu được cập nhật tại một trường dữ liệu của một CSDL thành phần sẽ tự động liên kết với trường cùng chức năng thông tin của CSDL thành phần khác để lấy dữ liệu phù hợp có sẵn hoặc cập nhật mới dữ liệu. Ví dụ, khi cập nhật dữ liệu vào trường “Cơ quan chủ trì nhiệm vụ” (có liên kết) của CSDL nhiệm vụ KH&CN, hệ thống sẽ tự động mở dữ liệu của trường “Tên tổ chức” của CSDL tổ chức KH&CN. Nếu tên cơ quan chủ trì nhiệm vụ đã có trong trường này, chỉ cần một động tác nháy chuột, tên cơ quan sẽ được cập nhật vào trường “Cơ quan chủ trì nhiệm vụ”. Trường hợp là tên mới, khi cập nhật vào trường “Cơ quan chủ trì nhiệm vụ” (CSDL nhiệm vụ KH&CN), dữ liệu sẽ tự động cập nhật vào các trường “Tên tổ chức” (CSDL tổ chức KH&CN), “Tên tổ chức nơi nhân lực làm việc” (CSDL nhân lực NC&PT), “Tổ chức KH&CN” (CSDL thống kê KH&CN), v.v...Mối liên kết trong chỉnh sửa dữ liệu cũng diễn ra tương tự;

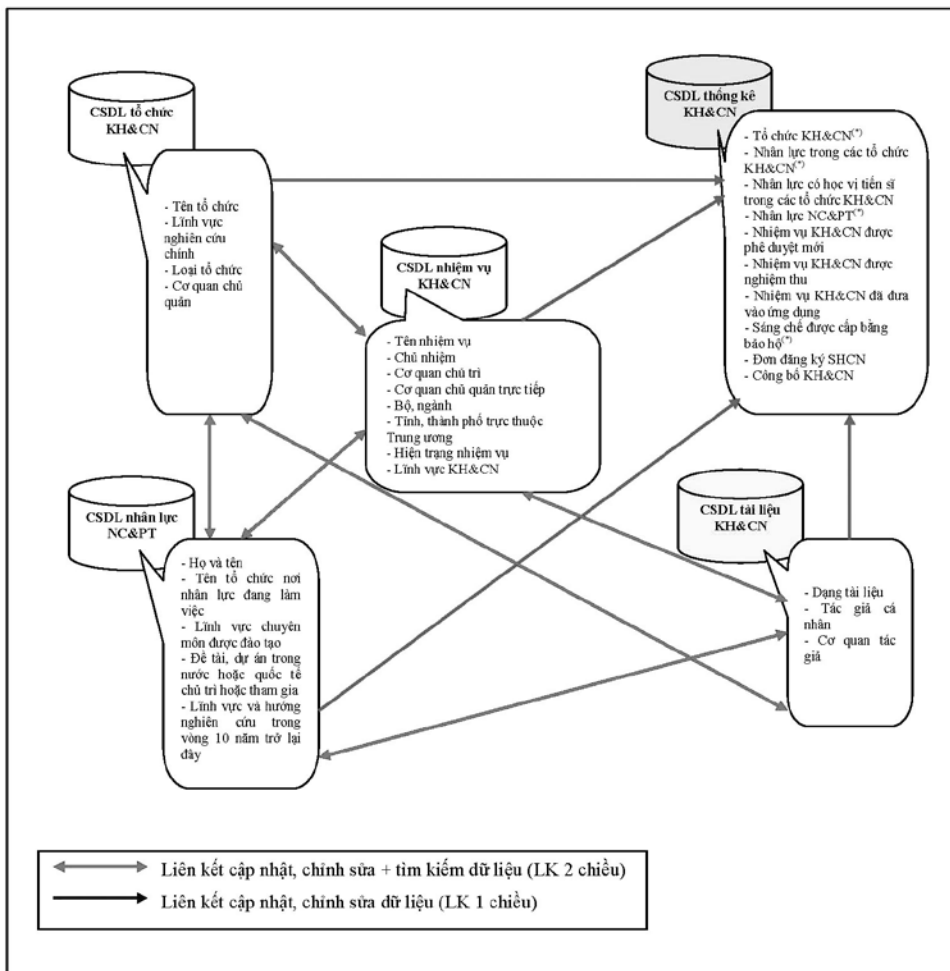
- *Liên kết trong tìm kiếm dữ liệu:* từ kết quả tìm tin của mỗi trường dữ liệu có liên kết bất kỳ của một CSDL thành phần bất kỳ, người dùng

tin có thể mở rộng phạm vi thông tin để có được thông tin một cách hệ thống bằng cách sử dụng đường kết nối đến các trường khác của các CSDL thành phần khác và ngược lại. Ví dụ, khi tìm nhân lực NC&PT, trường dữ liệu “Họ và tên” gọi mối liên tưởng đến “Tên tổ chức” nơi nhân lực làm việc (CSDL tổ chức KH&CN), “Tên nhiệm vụ” mà nhân lực đã hoặc đang tham gia (CSDL nhiệm vụ KH&CN). Một ví dụ khác, khi tìm thông tin về “Bằng sáng chế” (CSDL nhân lực NC&PT), người dùng tin có thể được cung cấp thông tin đầy đủ về nguồn gốc của sáng chế từ trường dữ liệu “Tên nhiệm vụ” (CSDL nhiệm vụ KH&CN) hoặc trường dữ liệu “Tên tổ chức” (CSDL tổ chức KH&CN) để có thêm thông tin về cơ quan chủ trì đề tài, dự án có

sáng chế nói trên, v.v...

CSDL thành phần có các trường thực hiện cả hai phương thức liên kết nói trên là các CSDL: tổ chức KH&CN, nhân lực NC&PT, nhiệm vụ KH&CN và tài liệu KH&CN. Có thể gọi liên kết này là liên kết hai chiều. Do đặc điểm CSDL thống kê KH&CN là dữ liệu trong CSDL này cơ bản là số liệu thống kê tổng hợp nên chúng tôi đề xuất chỉ áp dụng phương thức liên kết cập nhật, chỉnh sửa dữ liệu với các CSDL thành phần khác. Liên kết này có thể gọi là liên kết một chiều.

Hình 1 dưới đây là ví dụ về mối liên kết giữa một số trường dữ liệu của các CSDL thành phần thuộc CSDL quốc gia về KH&CN.



Hình 1. Ví dụ về mối liên kết giữa một số trường dữ liệu của các CSDL thành phần thuộc CSDL quốc gia về KH&CN

Đặc điểm nói trên nếu được chú trọng và có biện pháp giải quyết tốt thì không những tiết kiệm được thời gian và công sức cập nhật, chỉnh sửa, bổ sung dữ liệu mà còn nâng cao hiệu quả khai thác, sử dụng CSDL trên cơ sở bảo đảm cung cấp thông tin một cách hệ thống cho người dùng tin của CSDL quốc gia về KH&CN.

3. Yêu cầu đối với phần mềm quản trị cơ sở dữ liệu quốc gia về khoa học và công nghệ- tiếp cận từ đặc điểm cấu trúc dữ liệu

Căn cứ vào quy mô và tính chất dữ liệu của CSDL quốc gia về KH&CN, có nhiều yêu cầu đặt ra đối với phần mềm quản trị CSDL quốc gia về KH&CN. Ở đây, chúng tôi chỉ nêu một số yêu cầu xuất phát từ những đặc điểm chủ yếu về cấu trúc dữ liệu của CSDL đã trình bày ở trên.

Thứ nhất, phải đủ mạnh để quản lý được một hệ thống thông tin KH&CN lớn về số lượng trường dữ liệu và phong phú, đa dạng về nội dung thông tin như CSDL quốc gia về KH&CN;

Thứ hai, phải giải quyết tốt nhất hai trường hợp trùng lặp về trường dữ liệu. Trước hết, cần xác định, liệt kê đầy đủ các trường dữ liệu có trùng lặp và trường hợp trùng lặp (trùng lặp cả tên gọi và nội dung thông tin hay chỉ trùng lặp nội dung thông tin). Tiếp theo, cần có biện pháp giải quyết từng trường hợp, sao cho có thể tiết kiệm triệt để thời gian và công sức cập nhật, bổ sung, chỉnh sửa dữ liệu;

Thứ ba, phải giải quyết có hiệu quả các bài toán liên quan đến mối liên kết thông tin giữa một số trường dữ liệu trong mỗi CSDL thành phần cũng như trong tổng thể CSDL quốc gia về KH&CN. Để làm được việc này, cần xem xét, lập danh sách các trường dữ liệu có liên kết thông tin, xác định tính chất liên kết và giải pháp thực hiện, sao cho có thể tiết kiệm nhiều nhất thời gian và công sức cập nhật, bổ sung, chỉnh sửa dữ liệu. Đồng thời bảo đảm cung cấp thông tin một cách hệ thống cho người dùng tin của CSDL quốc gia về KH&CN.

Thứ tư, phải áp dụng các tiêu chuẩn và quy chuẩn kỹ thuật hiện đại, tiên tiến khi giải quyết

các vấn đề về cấu trúc dữ liệu CSDL quốc gia về KH&CN. Ưu tiên áp dụng các tiêu chuẩn, quy chuẩn kỹ thuật có hiệu lực bắt buộc áp dụng nhằm đảm bảo sự tương hợp và bảo đảm an toàn an ninh cho CSDL.

4. Kết luận

Với một hệ thống thông tin KH&CN to lớn và hiện đại cả về cấu trúc và đối tượng bao quát như CSDL quốc gia về KH&CN, có rất nhiều yêu cầu đặt ra đối với phần mềm quản trị CSDL. Yêu cầu giải quyết những vấn đề nảy sinh từ đặc điểm cấu trúc dữ liệu của CSDL chỉ là một phần rất nhỏ bên cạnh các yêu cầu khác đối với phần mềm, như: yêu cầu về năng lực quản trị hệ thống; yêu cầu về bảo đảm thực hiện và kiểm soát, giám sát thực hiện hoạt động cập nhật, truy cập, khai thác CSDL; yêu cầu về bảo đảm an toàn dữ liệu, an ninh hệ thống mạng; v.v... Tất cả những yêu cầu đối với phần mềm quản trị CSDL quốc gia về KH&CN nếu được đáp ứng sẽ góp phần thực hiện mục tiêu tạo lập một hệ thống thông tin tích hợp về KH&CN, đảm bảo phục vụ thông tin cho hoạt động quản lý nhà nước và hội nhập quốc tế về KH&CN của đất nước.

TÀI LIỆU THAM KHẢO

1. Nghị định số 11/2014-NĐ-CP ngày 18/02/2014 của Chính phủ về hoạt động thông tin KH&CN
2. Thông tư số 14/2014/TT-BKHCN của Bộ KH&CN quy định về việc thu thập, đăng ký, lưu giữ và công bố thông tin về nhiệm vụ KH&CN, ban hành ngày 11/6/2014
3. Lê Xuân Định và các cộng sự (2013). Nghiên cứu xây dựng khung và hệ tiêu chí phát triển CSDL KH&CN quốc tế cho hoạt động hội nhập quốc tế về KH&CN của Việt Nam/BCTH kết quả đề tài nghiên cứu KH&CN cấp nhà nước, mã số KX06.02/11-15.-Cơ quan chủ trì: Cục TTKHCNQG.-Năm hoàn thành: 2013
4. Nguyễn Hồng Hạnh và các cộng sự (2016). Nghiên cứu đề xuất cấu trúc, tiêu chí dữ liệu CSDL quốc gia về KH&CN, Dự thảo BCTH kết quả nghiên cứu KH&CN cấp bộ.-Cơ quan chủ trì: Trung tâm TTLKH&CNQG.-N: 2016.

(Ngày Tòa soạn nhận được bài: 10-4-2017; Ngày phản biện đánh giá: 28-5-2017; Ngày chấp nhận đăng: 28-6-2017).