



BỘ KHOA HỌC CÔNG NGHỆ VÀ MÔI TRƯỜNG
TRUNG TÂM THÔNG TIN TƯ LIỆU
KHOA HỌC & CÔNG NGHỆ QUỐC GIA

PHAN HUY QUẾ

MÔ TẢ NỘI DUNG TÀI LIỆU BẰNG TỪ KHÓA

(ĐỊNH TỪ KHÓA TÀI LIỆU)

TÀI LIỆU HƯỚNG DẪN

HÀ NỘI - 2001



BỘ KHOA HỌC CÔNG NGHỆ VÀ MÔI TRƯỜNG
TRUNG TÂM THÔNG TIN TƯ LIỆU
KHOA HỌC & CÔNG NGHỆ QUỐC GIA

PHAN HUY QUẾ

MÔ TẢ NỘI DUNG TÀI LIỆU BẰNG TỪ KHÓA

(ĐỊNH TỪ KHÓA TÀI LIỆU)

TÀI LIỆU HƯỚNG DẪN

HÀ NỘI - 2001

LỜI NÓI ĐẦU

Mô tả nội dung tài liệu bằng từ khóa (định từ khóa tài liệu) là một trong những công đoạn của qui trình xử lý thông tin, là việc làm thường xuyên hiện nay của các cơ quan thông tin tư liệu khi tiến hành xây dựng cơ sở dữ liệu tư liệu. Chất lượng định từ khóa tài liệu có ảnh hưởng rất lớn đến chất lượng phục vụ thông tin, đặc biệt là việc tìm kiếm bằng ngôn ngữ từ khóa những thông tin phù hợp với yêu cầu của người dùng. Tuy nhiên, định từ khóa tài liệu cũng là một trong những công việc phức tạp nhất của quy trình xử lý thông tin, xuất phát từ tính chất phức tạp của ngôn ngữ, tính đa dạng của nội dung tài liệu v.v... Do đó, rất cần có một tài liệu hướng dẫn mang tính phương pháp luận quy định những yêu cầu cơ bản đối với từ khóa và hướng dẫn cụ thể phương pháp định từ khóa tài liệu, nhằm đảm bảo tính thống nhất về từ khóa trong Hệ thống thông tin KH&CN nói chung và trong mỗi hệ thống tìm tin tư liệu nói riêng. Rất tiếc, từ trước đến nay, chúng ta chưa có được một tài liệu như vậy nên việc định từ khóa có lúc chưa đạt được kết quả tương xứng với vai trò và nhiệm vụ của nó trong hệ thống tìm tin tư liệu.

Với mong muốn nâng cao chất lượng và đảm bảo tính thống nhất của việc định từ khóa tài liệu trong các

cơ quan thông tin tư liệu của Hệ thống thông tin tư liệu KH&CN quốc gia, được sự chỉ đạo và giúp đỡ của lãnh đạo Trung tâm Thông tin Tư liệu KH&CN quốc gia , chúng tôi mạnh dạn biên soạn cuốn tài liệu hướng dẫn “Mô tả nội dung tài liệu bằng từ khóa” nhằm đóng góp một tài liệu tham khảo bước đầu cho những người trực tiếp định từ khóa trong các cơ quan thông tin tư liệu.

Tác giả trân trọng bày tỏ lòng biết ơn sự giúp đỡ nhiệt tình của nhiều đồng nghiệp trong và ngoài Trung tâm Thông tin Tư liệu KH&CN quốc gia, đặc biệt với TS Nguyễn Thu Thảo, người đã góp nhiều ý kiến quý báu trong quá trình biên soạn tài liệu này.

Do đây là tài liệu hướng dẫn đầu tiên về định từ khóa nên chắc chắn không tránh khỏi những khiếm khuyết. Tác giả mong được sự lượng thứ và góp ý của đồng nghiệp. Xin chân thành cảm ơn.

Hà Nội, ngày 28 tháng 02 năm 2001

TÁC GIẢ

1. NHỮNG VẤN ĐỀ CHUNG

1.1. Một số định nghĩa

a. Từ khóa: Từ hoặc cụm từ ổn định, đơn nghĩa được sử dụng để mô tả nội dung chính của tài liệu và để tìm tin trong hệ thống tìm tin tư liệu.

b. Mô tả nội dung tài liệu bằng từ khóa (Định từ khóa): Quá trình phân tích nội dung tài liệu và chọn từ khóa để mô tả nội dung chính của tài liệu.

c. Ngôn ngữ từ khóa : Ngôn ngữ tìm tin hậu kết hợp, được sử dụng trong quá trình xử lý nội dung tài liệu và yêu cầu tin.

d. Ngôn ngữ tìm tin: Ngôn ngữ nhân tạo được sử dụng trong quá trình xử lý tài liệu và tìm tin trong hệ thống tìm tin tư liệu.

e. Xử lý thông tin: Tập hợp các công đoạn nhằm biến đổi thông tin thu thập được thành các dạng thể hiện mới.

f. Tài liệu: Vật mang tin, được xem như một đối tượng xử lý trong quá trình xử lý thông tin.

g. Hệ thống tìm tin tư liệu: Hệ thống tìm tin được xây dựng để tìm và cung cấp thông tin về tài liệu hoặc bản thân tài liệu.

h. Cơ sở dữ liệu tư liệu: Tập hợp có cấu trúc các bản mô tả tài liệu và/hoặc bản thân tài liệu, được lưu trữ và tìm kiếm trên máy tính.

i. Biểu ghi: Bản mô tả các đặc tính của tài liệu, hình thành trong quá trình xử lý tài liệu, được lưu trữ trong cơ sở dữ liệu tư liệu.

j. Khái niệm: Đơn vị tư tưởng thiết lập qua trừu tượng hóa trên cơ sở khái quát những mặt, những dấu hiệu cơ bản của sự vật, hiện tượng, quá trình...trong thế giới khách quan.

k. Hệ khái niệm: Tập hợp kết cấu nhiều khái niệm xây dựng trên cơ sở các liên hệ giữa các khái niệm đó, và trong đó mỗi khái niệm được xác định bởi vị trí của nó trong tập hợp.

l. Đối tượng nghiên cứu của tài liệu: Khái niệm hoặc sự vật cụ thể được xem xét trong nội dung tài liệu.

m. Phương diện (khía cạnh) nghiên cứu của đối tượng: Một mặt nào đó của đối tượng của tài liệu được nghiên cứu, xem xét trong nội dung tài liệu.

1.2. Vai trò của từ khóa trong hệ thống tìm tin tư liệu

a. Định từ khóa và quy trình xử lý thông tin

Những công đoạn chính của quy trình xử lý thông tin có thể được sắp xếp theo thứ tự thực hiện như sau:

- mô tả thư mục;
- phân loại;
- chú giải/tóm tắt;
- định từ khóa;

- tổng quan.

Người ta có thể áp dụng tất cả hoặc một số các công đoạn kể trên đối với một tài liệu khoa học. Ví dụ: có những tài liệu được xử lý từ công đoạn mô tả thư mục đến công đoạn tổng quan; nhưng cũng có tài liệu chỉ được xử lý ở một số công đoạn cần thiết như mô tả thư mục, phân loại và định từ khóa; v.v... Việc áp dụng toàn bộ hay một số công đoạn của quy trình phụ thuộc vào các yếu tố như: tính chất tài liệu, mục tiêu/nhiệm vụ của cơ quan xử lý thông tin, đặc điểm của hệ thống thông tin... Nhưng nếu áp dụng toàn bộ các công đoạn của quy trình để xử lý một tài liệu khoa học, thì việc thực hiện mỗi công đoạn theo thứ tự đã sắp xếp ở trên là một trong những yếu tố đảm bảo hiệu quả của hoạt động xử lý thông tin, nghĩa là nếu công đoạn định từ khóa được thực hiện sau công đoạn làm chú giải/tóm tắt tài liệu thì hiệu quả sẽ được đảm bảo hơn.

b. Từ khóa - đơn vị từ vựng thông báo nội dung tài liệu khoa học

Quá trình xử lý thông tin là quá trình biến đổi thông tin thành các dạng thể hiện mới. Do đó, những thông tin của nội dung tài liệu được "cô đặc" lại về khối lượng và thể hiện qua kết quả của mỗi công đoạn của quy trình. Kết quả của công đoạn định từ khóa là một tập hợp các thuật ngữ được chọn ra từ nội dung tài liệu, đại diện cho những thông tin cơ bản của nội dung tài liệu. Tập hợp thuật ngữ này thông báo cho người dùng

tin về những vấn đề cơ bản được đề cập trong nội dung tài liệu.

c. Từ khóa - điểm truy nhập nội dung tài liệu bằng từ ngữ

Một trong những mục tiêu quan trọng của xử lý thông tin là thiết lập các điểm truy nhập mà theo đó, người dùng tin có thể nhanh chóng và dễ dàng tìm đến và khai thác những thông tin phù hợp với yêu cầu của họ. Ví dụ: mục tiêu của công đoạn mô tả thư mục là thiết lập điểm truy nhập về tên tài liệu, tác giả tài liệu và các yếu tố liên quan đến xuất bản tài liệu... Mục tiêu của công đoạn định từ khóa là thiết lập điểm truy nhập nội dung tài liệu bằng từ ngữ. Từ điểm truy nhập này, người dùng tin có thể tiếp cận và khai thác một/một số tài liệu có nội dung được thể hiện bằng tập hợp các từ ngữ phù hợp với yêu cầu của họ. Đây là vai trò quan trọng nhất của từ khóa trong hệ thống tìm tin tư liệu.

1.3. Phân loại từ khóa

Căn cứ vào đặc điểm từ vựng và chức năng, người ta phân loại từ khóa theo một số tiêu chí cơ bản như sau:

a. Phân loại theo tiêu chí từ vựng

Theo tiêu chí này, từ khóa có thể có các loại sau:

- Từ khóa là danh từ:

- Ví dụ: + *sinh viên* (danh từ chung chỉ nhóm người);
+ *cá voi* (danh từ chung chỉ động vật);
+ *chò chỉ* (danh từ chung chỉ thực vật);
+ *máy tính* (danh từ chung chỉ đồ vật);

- + *Việt Nam* (danh từ riêng chỉ địa danh);
- + *Hồ Chí Minh* (danh từ riêng chỉ người);
- + v.v...

- Từ khóa là cụm từ gồm danh từ + tính từ (hoặc danh từ làm nhiệm vụ bổ nghĩa):

- Ví dụ: + *máy bay phản lực*;
+ *điều tra cơ bản*;
+ *đất bazan*;
+ v.v...

- Từ khóa là động từ, được hiểu như danh từ:

- Ví dụ: + *sản xuất*;
+ *hoàn thiện*;
+ *phòng chống*;
+ v.v...

- Từ khóa là cụm từ gồm động từ + danh từ

- Ví dụ: + *gia công bề mặt* ;
+ *trộn bê tông*;
+ *phân tích hệ thống*;
+ v.v...

b. Phân loại theo tiêu chí nội dung thông tin

Theo tiêu chí này, từ khóa được phân thành 2 loại: từ khóa đối tượng và từ khóa phương diện dựa trên cơ sở từ khóa mang thông tin gì về nội dung tài liệu.

Theo phương pháp luận xử lý thông tin, khi phân tích nội dung tài liệu khoa học người ta thường tập trung làm rõ 2 yếu tố sau đây:

b.1. Đối tượng nghiên cứu của tài liệu: tức là xác định tài liệu đề cập đến đối tượng gì? Đối tượng ở đây có thể là hữu hình (máy bay, bộ đội, cây sả...) hoặc vô hình (phương pháp, quy trình, từ trường, không khí...).

Một tài liệu khoa học có thể có một hoặc nhiều hơn một đối tượng nghiên cứu.

Ví dụ:

- *tài liệu có một đối tượng:* **Thiết kế nhà nổi.** (Trong ví dụ này, nhà nổi là đối tượng nghiên cứu của tài liệu);

- *tài liệu có hơn một đối tượng:* **Đào tạo cán bộ thông tin và cán bộ thư viện ở Việt Nam** (Trong ví dụ này, cán bộ thông tin , cán bộ thư viện là 2 đối tượng nghiên cứu của tài liệu).

Các đối tượng trong các ví dụ trên là đối tượng chính mà nội dung tài liệu đề cập đến và được coi là **đối tượng bậc 1**. Trong cùng một tài liệu, ngoài đối tượng bậc 1, còn có các *đối tượng bậc 2*. **Đối tượng bậc 2** có thể là:

- một đối tượng độc lập với đối tượng bậc 1 nhưng được đề cập trong mối quan hệ với đối tượng bậc 1. Ví dụ: trong một tài liệu khoa học có chủ đề **"Ứng dụng kỹ thuật số trong công nghệ chụp ảnh"**. Ở đây, kỹ thuật số là đối tượng bậc 1. Nhưng *kỹ thuật số* được

xem xét trong mối quan hệ với một đối tượng khác là *công nghệ chụp ảnh* (đối tượng bao trùm). Trong trường hợp này, công nghệ chụp ảnh là đối tượng bậc 2 của tài liệu;

- một thành phần/bộ phận của đối tượng bậc 1 nhưng độc lập về khái niệm và bản chất. Ví dụ: trong một tài liệu khoa học có chủ đề: “**phần cứng máy tính**”. Ở đây, phần cứng máy tính là đối tượng bậc 1. Nhưng trong nội dung tài liệu, tác giả tài liệu chỉ xem xét 3 thành phần chủ yếu của phần cứng máy tính là: **bộ xử lý trung tâm (CPU); màn hình; bàn phím**. Trong trường hợp này, bộ xử lý trung tâm (CPU); màn hình; bàn phím là các đối tượng bậc 2.

Đối tượng bậc 2 có thể không nằm trong nhan đề của tài liệu khoa học mà chỉ được đề cập trong khi trình bày nội dung tài liệu.

Những từ khóa phản ánh các đối tượng bậc 1 và đối tượng bậc 2 của nội dung tài liệu được gọi là **từ khóa đối tượng**.

b.2. Phương diện của đối tượng

Phương diện (khía cạnh) của đối tượng là cách thức tiếp cận đối tượng của nội dung tài liệu khoa học. Phương diện của đối tượng cũng được phân loại thành 2 bậc: *phương diện bậc 1 và phương diện bậc 2*.

Phương diện bậc 1 đi liền với đối tượng, bổ nghĩa trực tiếp cho đối tượng.

Ví dụ: *Bảo dưỡng động cơ máy bay*;

Phát triển mô hình VAC

Trong các ví dụ trên, bảo dưỡng, phát triển là phương diện bậc 1 của các đối tượng động cơ máy bay và mô hình VAC.

Phương diện bậc 2 thường không bổ nghĩa trực tiếp cho đối tượng mà đi liền với phương diện bậc 1 và bổ nghĩa cho phương diện bậc 1.

Ví dụ: *Nghiên cứu sản xuất sợi cách nhiệt.*

Phương pháp xây dựng chỉ tiêu thống kê khoa học công nghệ.

Trong các ví dụ trên, nghiên cứu, phương pháp là các phương diện bậc 2 vì chúng bổ nghĩa trực tiếp cho các phương diện bậc 1 là sản xuất và xây dựng.

Những từ khóa phản ánh các phương diện bậc 1 và bậc 2 được gọi là từ khóa **phương diện**.

c. Phân biệt theo tiêu chí tải trọng thông tin trong nội dung tài liệu

Theo tiêu chí này, từ khóa được chia thành 2 loại: từ khóa chính và từ khóa phụ.

c.1. Từ khóa chính bao gồm:

- toàn bộ các từ khóa mô tả đối tượng bậc 1;

- từ khóa mô tả đối tượng bậc 2: chỉ nên chọn những đối tượng có trọng số cao về khối lượng thông tin trong nội dung tài liệu (xét trong mối quan hệ với đối tượng bậc 1).

c.2. Từ khóa phụ bao gồm:

- từ khóa mô tả đối tượng bậc 2 không có trọng số cao về thông tin trong nội dung tài liệu ;
- từ khóa mô tả phương diện bậc 1 ;
- từ khóa mô tả phương diện bậc 2.

Ngoài các tiêu chí phân loại nói trên, tùy theo đặc điểm của hệ thống thông tin, người ta còn chia từ khóa thành các loại sau:

- từ khóa địa lý: từ khóa chính (hoặc phụ) chỉ địa danh;
- từ khóa nhân vật: từ khóa chính chỉ tên người;
- từ khóa sinh vật: từ khóa chính chỉ tên sinh vật bằng tiếng latin;
- từ khóa thời gian: từ khóa chính (hoặc phụ) phản ánh khoảng thời gian của vấn đề được đề cập trong nội dung tài liệu;
- v.v...

1.4. Một số yêu cầu chung

Một trong những đặc điểm của việc định từ khóa tài liệu khoa học là dùng các từ ngữ ổn định để mô tả nội dung tài liệu, tức là hình thức hóa nội dung tài liệu bằng các từ ngữ và sử dụng các từ ngữ này như một dấu hiệu hình thức để tìm tài liệu theo nội dung. Đặc điểm này đặc biệt quan trọng đối với hệ thống thông tin tự động hóa. Trong tìm tin tự động, quá trình tìm

tin là quá trình hệ thống so sánh hình thức các từ khóa trong lệnh tìm (từ khóa được rút ra từ yêu cầu tin) với các từ khóa trong mẫu tìm (từ khóa được rút ra từ nội dung tài liệu). Do đó, việc có tìm và tìm được đầy đủ được những tài liệu phù hợp với yêu cầu tin hay không phụ thuộc vào một số yếu tố chính sau đây:

- độ chính xác và đầy đủ của từ khóa khi xử lý nội dung tài liệu;

- độ chính xác và đầy đủ của từ khóa khi xử lý yêu cầu tin và xây dựng các biểu thức tìm;

- tính thống nhất về hình thức giữa các từ khóa trong mẫu tìm và lệnh tìm.

Vì vậy, cần có những yêu cầu cụ thể đối với từ khóa trong quá trình xử lý từ vựng để bảo đảm độ đầy đủ, chính xác và tính thống nhất của từ khóa như đã nêu ở trên.

1.4.1. Yêu cầu đối với từ khóa

1.4.1.1. Yêu cầu về nội dung từ khóa

a. Thông dụng, đúng đắn theo thuật ngữ khoa học

a.1. Nội dung yêu cầu: từ khóa phải là từ khoa học thông dụng trong lĩnh vực mà nội dung tài liệu đề cập, không sử dụng khẩu ngữ và các từ nghĩa bóng.

a.2. Biện pháp đảm bảo:

- sử dụng các từ điển chuyên ngành, hoặc

- nâng cao kiến thức chuyên ngành cho cán bộ xử lý thông tin, hoặc

- lấy ý kiến chuyên gia về lĩnh vực chuyên ngành.

Ví dụ:

<u>Không nên dùng:</u>	<u>Nên dùng:</u>
<i>Hệ thống mặt trời</i>	<i>Hệ Mặt trời</i>
<i>Ban bố</i>	<i>Ban hành</i>
<i>Biên giới quốc gia trên không</i>	<i>Không phận</i>
<i>Hóa</i>	<i>Hóa học</i>
<i>Cái chết trắng</i>	<i>Hêrôin</i>

b. Súc tích

b.1. Nội dung yêu cầu: từ khóa phải thể hiện nội dung thông tin dưới hình thức ngắn gọn nhất. Yêu cầu này nhằm định hướng vào việc chọn lựa những từ thực sự có nội dung thông tin và loại bỏ những từ không có ích cho việc tra cứu.

b.2. Biện pháp đảm bảo: nén từ, bằng các cách sau đây:

- giảm lược các hư từ như: sự, những, các, liên từ, giới từ (trong những trường hợp có thể);

Ví dụ:

<u>Không nên dùng:</u>	<u>Nên dùng:</u>
<i>Sự lão hóa</i>	<i>Lão hóa</i>
<i>Những thiết bị chuyên dụng</i>	<i>Thiết bị chuyên dụng</i>

<i>Các quốc gia phát triển</i>	<i>Quốc gia phát triển</i>
<i>Ngân hàng của tư nhân</i>	<i>Ngân hàng tư nhân</i>

- giản lược các thực từ không làm rõ thêm ý nghĩa cho từ chính (trong những trường hợp có thể);

Ví dụ:

<u>Không nên dùng:</u>	<u>Nên dùng:</u>
<i>Bệnh AIDS</i>	<i>AIDS</i>
<i>Hiện tượng nhật thực</i>	<i>Nhật thực</i>
<i>Đường xích đạo</i>	<i>Xích đạo</i>
<i>Chi bộ Đảng</i>	<i>Chi bộ</i>

- giản lược các từ có khả năng làm tản mạn vốn từ vựng.

Ví dụ:

<u>Không nên dùng:</u>	<u>Nên dùng:</u>
<i>Ngành thuỷ sản</i>	<i>Thủy sản</i>
<i>Tuyển chọn công chức</i>	<i>Tuyển công chức</i>
<i>Khu vực bảo tồn thiên nhiên</i>	<i>Khu bảo tồn thiên nhiên</i>
<i>Nền văn hóa</i>	<i>Văn hóa</i>
<i>Cây lúa</i>	<i>Lúa</i>

c. Ngắn gọn

c.1. *Nội dung yêu cầu*: tách các khái niệm phức tạp thành các khái niệm đơn giản nhất có thể. Yêu cầu này nhằm tạo điều kiện thuận lợi cho việc đảm bảo tính hậu kết hợp của ngôn ngữ từ khóa.

c.2. *Biện pháp đảm bảo*: tách từ, có hai cách tách từ như sau:

- tách theo cấu trúc cú pháp

Ví dụ:

Cấu trúc cú pháp	Mẫu cụm từ	Kết quả tách
Cụm từ gồm hành động hướng ngoại và đối tượng của hành động	<i>Chế tạo máy công cụ</i>	1. <i>Chế tạo</i> 2. <i>Máy công cụ</i>
Cụm từ gồm hành động hướng nội và chủ thể của hành động	<i>Sự lão hóa của vật liệu nhựa</i>	1. <i>Lão hóa</i> 2. <i>Vật liệu nhựa</i>
Cụm từ gồm đối tượng và tính chất (hoặc chỉ tiêu, thông số...) của đối tượng	<i>Hàm lượng thủy ngân</i>	1. <i>Hàm lượng</i> 2. <i>Thủy ngân</i>
Cụm từ chỉ chủ thể và bộ phận của chủ thể	<i>Động cơ ô tô</i>	1. <i>Động cơ</i> 2. <i>Ô tô</i>

- Tách theo cấu trúc ngữ nghĩa: Người ta phân tích cấu trúc ngữ nghĩa của cụm từ và tách chúng theo 2 trường hợp chính căn cứ vào quan hệ ngữ nghĩa của chúng như sau:

+ trường hợp 1: từ bổ nghĩa phía sau bổ nghĩa cho từ đứng ngay trước nó.

Ví dụ:

Mẫu cụm từ	Quan hệ ngữ nghĩa	Kết quả tách
Tạo màng sinh học	Từ "màng" bổ nghĩa cho từ "tạo", từ "sinh học" bổ nghĩa cho từ "màng"	1. Tạo màng 2. Màng sinh học
Cải cách hành chính địa phương	Từ "hành chính" bổ nghĩa cho từ "cải cách", từ "địa phương" bổ nghĩa cho từ "hành chính"	1. Cải cách hành chính 2. Hành chính địa phương
Chụp ảnh nghệ thuật	Từ "ảnh" bổ nghĩa cho từ "chụp", từ "nghệ thuật" bổ nghĩa cho từ "ảnh"	1. Chụp ảnh 2. ảnh nghệ thuật

+ trường hợp 2: các từ phía sau đều bổ nghĩa cho từ đứng thứ nhất

Ví dụ:

Mẫu cụm từ	Quan hệ ngữ nghĩa	Kết quả tách
Đào tạo cao học từ xa	Từ "cao học" bổ nghĩa cho từ "đào tạo", từ "từ xa" cũng bổ nghĩa cho từ "đào tạo"	1. Đào tạo cao học 2. Đào tạo từ xa
Thép cacbon bền nhiệt	Từ "cacbon" bổ nghĩa cho từ "thép", từ "bền nhiệt" cũng bổ nghĩa cho từ "thép"	1. Thép cacbon 2. Thép bền nhiệt
Chiến tranh du kích cục bộ	Từ "du kích" bổ nghĩa cho từ "chiến tranh", từ "cục bộ" cũng bổ nghĩa cho từ "chiến tranh"	Chiến tranh du kích Chiến tranh cục bộ

* **Chú ý:** Không nên tách các cụm từ ổn định trong những trường hợp sau:

+ Những cụm từ khó xác định là từ hay là ngữ ở chỗ chúng có thể bị biến đổi ý nghĩa hoặc trở nên đa nghĩa nếu bị tách thành các từ đơn.

Ví dụ: *Thượng tầng kiến trúc; Xóa mù chữ; Cấy mô; Dạng tài liệu; Chi phí lao động; Chiến lược công nghệ; v.v...*

+ Những cụm từ định hướng nhiệm vụ ưu tiên của hệ thống cụ thể (thường dựa theo sự trợ giúp của phương tiện kiểm soát từ)

Ví dụ: *Chỉ dẫn thư mục; Chỉ số axit; Kiểm soát đầu vào; v.v...;*

+ Những từ ghép được dùng với độ ổn định cao trong thực tế của chuyên ngành cụ thể của CSDL, bao gồm những từ gần giống nhau và có thể được quan tâm đồng thời với nhau trong chuyên ngành đó.

Ví dụ: *Cát sỏi; Cầu đường; Hoa quả; Quần áo; Bát đĩa; Nuôi trồng; Chai lọ; v.v...;*

+ Tên gọi các lĩnh vực hoạt động, các ngành, chuyên ngành và bộ môn khoa học thuộc diện bao quát đề tài của CSDL cụ thể.

Ví dụ: *Bảo hộ lao động; Kỹ thuật điện tử; Vật lý bán dẫn; Nghiên cứu lịch sử Đảng; Bảo vệ môi trường; v.v...;*

+ Các thuật ngữ phái sinh thuộc hệ thống khái niệm cơ bản của các thuật ngữ gốc thuộc lĩnh vực quan tâm, ưu tiên trong diện bao quát đề tài.

Ví dụ:

Thuật ngữ gốc	Các thuật ngữ phái sinh
<i>Vĩ sinh học</i>	<i>Vĩ sinh học công cộng</i> <i>Vĩ sinh học công nghiệp</i> <i>Vĩ sinh học kỹ thuật</i> <i>Vĩ sinh học nông nghiệp</i> <i>Vĩ sinh học nước.</i> <i>v.v...</i>

+ Tên gọi các định lý, định luật, lý thuyết, phương pháp thuộc hệ khái niệm cơ bản của diện bao quát đề tài của CSDL cụ thể. Ví dụ: *Định lý đảo; Định lý Betti; Định luật bảo toàn năng lượng; Định luật khuếch tán Fick; Lý thuyết nguyên tử; Lý thuyết Huckel; Phương pháp đồ thị; Phương pháp Pick Kirinski; v.v...*;

+ Tên gọi các nhóm đối tượng là máy móc, thiết bị, thuốc, vật liệu đi kèm với chức năng, hoặc tính chất, hoặc phương thức hoạt động. Ví dụ: *Máy X quang; Máy sao chụp; Dụng cụ cắt gọt; Dụng cụ đo chất khí; Động cơ Stirling; Thiết bị siêu cao tần; Thuốc nhuộm bazo; Chất lỏng phát quang; Chất kích nổ; v.v...*;

+ Một số ít các cụm từ được sử dụng với độ ổn định cao trong các ngành ưu tiên của CSDL cụ thể. Ví dụ: *Đa*

nguyên chính trị; Đa tạp compact; Công nghệ linh hoạt; v.v...

d. Đơn nghĩa

d.1. Nội dung yêu cầu: một từ khóa chỉ mang một nghĩa duy nhất và một khái niệm (hoặc đối tượng cụ thể) chỉ được mô tả bằng một từ khóa duy nhất. Yêu cầu này nhằm khắc phục các hiện tượng đa nghĩa và đồng nghĩa trong khi xử lý từ vựng.

d.2. Biện pháp đảm bảo:

- Đối với trường hợp từ đa nghĩa: nếu một từ khóa mang nhiều nghĩa khác nhau thì phải biến nó thành từ đơn nghĩa bằng cách sử dụng cụm từ thay cho từ đơn (không thực hiện việc nén từ).

Ví dụ: không nên thực hiện thao tác nén các cụm từ sau đây:

- *Cây vôi voi;*
- *Cây đinh;*
- *Phép đối xứng;*
- *Đường thực phẩm;*
- *v.v...*

Nếu biện pháp nói trên làm cho từ khóa quá rườm rà thì có thể giữ nguyên từ khóa đa nghĩa nhưng phải áp dụng một số giải pháp khác trong phương tiện kiểm soát từ (ví dụ có các quy định cụ thể trong quy tắc lập mẫu tìm và lệnh tìm; các quy ước, quy định trong các phương tiện hỗ trợ như Từ điển từ khóa , Từ điển từ chuẩn...).

Nếu một từ khóa đa nghĩa có các nghĩa gần giống nhau (xét trên quan điểm tìm tin) thì có thể chấp nhận sử dụng.

Ví dụ: có thể giữ nguyên các từ khóa sau đây:

- *Giáo dục;*
- *Văn hóa;*
- *Lúa;*
- *Ngô;*
- *Cây.*
- *v.v...*

- Đối với trường hợp từ đồng nghĩa: sử dụng đồng thời các phương án đồng nghĩa trong một biểu ghi (nếu không có phương tiện kiểm soát từ vựng), hoặc sử dụng một từ trong số các từ đồng nghĩa được quy định trong các phương tiện kiểm soát từ vựng.

Ví dụ:

Từ đồng nghĩa	Biện pháp giải quyết
<i>Lợn</i> <i>Heo</i>	- <i>sử dụng cả 2 từ;</i> - <i>chỉ sử dụng 1 từ thống nhất theo phương tiện kiểm soát từ vựng.</i>
<i>Nhân quyền</i> <i>Quyền con người</i>	- <i>sử dụng cả 2 từ;</i> - <i>chỉ sử dụng 1 từ thống nhất theo phương tiện kiểm soát từ vựng.</i>

e. Chính xác, hiện đại

e.1. Nội dung yêu cầu: từ khóa phải phản ánh chính xác những khái niệm cơ bản của nội dung tài liệu, đồng thời phải là các thuật ngữ được dùng hiện tại thuộc các lĩnh vực chuyên ngành. Yêu cầu này giúp đảm bảo độ chính xác và tính thống nhất của từ khóa trong mẫu tìm và lệnh tìm.

e.2. Biện pháp đảm bảo:

- Sử dụng các phương tiện hỗ trợ như : TCVN (Tiêu chuẩn Việt nam); Từ điển chuyên ngành (xuất bản mới nhất); Tài liệu khoa học chuyên ngành (xuất bản mới nhất); ý kiến chuyên gia chuyên ngành;

- Trường hợp khái niệm trong nội dung tài liệu quá sâu so diện bao quát của hệ thống thông tin tư liệu, thì có thể thay thế hoặc bổ sung thêm khái niệm bao trùm ở mức gần nhất.

Ví dụ:

<u>Khái niệm trong nội dung tài liệu:</u>	<u>Từ khóa:</u>
<i>Xe gắn máy chạy bằng ắc quy khô</i>	<i>Xe gắn máy</i>
<i>Xử lý chất thải trong sản xuất nhựa</i>	<i>Xử lý chất thải</i>
<i>Chương trình phủ xanh đồi trọc</i>	<i>Trồng rừng</i>

f. Khách quan

f.1. Nội dung yêu cầu: từ khóa phải độc lập với văn cảnh của nội dung tài liệu gốc, tức là có thể hiểu được từ khóa mà không cần đọc nội dung tài liệu gốc. Đồng thời từ khóa không được mang sắc thái phê phán, đánh giá.

f.2. Biện pháp đảm bảo: bỏ các từ phụ thuộc vào văn cảnh nội dung tài liệu gốc hoặc những từ mang sắc thái phê phán, đánh giá.

Ví dụ:

Thuật ngữ phụ thuộc vào văn cảnh nội dung tài liệu hoặc mang sắc thái phê phán đánh giá:

Từ khóa:

80 bài toán mẫu

Bài toán mẫu

Hệ thống thông tin thích hợp

Hệ thống thông tin

Xi măng giá rẻ

Xi măng

2.4.1.2. Yêu cầu về hình thức từ khóa

Tất cả các yêu cầu về hình thức từ khóa dưới đây nhằm đảm bảo tính thống nhất về hình thức từ khóa, nâng cao hiệu quả tìm tin.

a. Yêu cầu về chính tả tiếng Việt

a.1. *Nội dung yêu cầu:* từ khóa nên viết theo cách viết thông dụng về chính tả tiếng Việt đối với vị trí dấu thanh, chữ y và i.

a.2. *Biện pháp đảm bảo:*

- tuân thủ các viết trong các tài liệu như: Từ điển tiếng Việt (xuất bản mới nhất); TCVN về phong chữ tiếng Việt...

b. Yêu cầu về danh từ chung có gốc tiếng nước ngoài

b.1. *Nội dung yêu cầu:* từ khóa là danh từ chung gốc tiếng nước ngoài phải tuân thủ cách viết và cách phiên chuyển thông dụng.

b.2. *Biện pháp đảm bảo:*

- Tên hóa chất và men sinh học: không mô tả các ký hiệu và công thức hóa học, chỉ mô tả theo tên thông dụng, viết theo TCVN 5529-1991.

Ví dụ: không dùng H_2SO_4 mà dùng *axit sunfuric*

- Tên thuốc: viết theo tên gốc do hãng sản xuất quy định;

Ví dụ: không viết *Pênixilin* mà viết *Penicilin*

không viết *Ký ninh* mà viết *Quinin*

- Tên sinh vật: viết tên Việt Nam và tên Latinh của sinh vật

Ví dụ: *Hà thủ ô đỏ* và *Fallopia multiflora*

Hạc đen và *Ciconia nigra*

- Từ đã được Việt hóa thông dụng: viết theo cách viết thông dụng.

Ví dụ: *ô tô; xi măng; bê tông...*

Tuy nhiên, trong thực tế hiện nay, khó xác định được cách viết nào là thông dụng nhất. Vì vậy, cần phải có những quy ước, quy định chặt chẽ, hoặc cao hơn là cần có những Tiêu chuẩn Việt Nam cho cách viết các loại từ này.

c. Yêu cầu về viết danh từ riêng

- Tên người có gốc chữ latin: sử dụng nguyên tên gốc (trừ những trường hợp tên các nhà triết học, chính trị học kinh điển đã được viết theo cách viết tiếng Việt thông dụng).

Ví dụ: *Bill Clinton; Al Gore; v.v...;*

Pla tôn; Hê ghen; Các Mác; Ăng ghen; Lê nin; v.v...

- Tên người có gốc chữ Kiril: phiên âm sang chữ Latinh

Ví dụ: viết là *Pushkin*

- Tên người có gốc chữ Trung Quốc: phiên theo âm Hán - Việt.

Ví dụ: *Mao Trạch Đông*

Lỗ Tấn

Tề Bạch Thạch

- Tên địa danh: viết theo các phương tiện kiểm soát từ vựng (các TCVN hoặc các tài liệu chuẩn khác);

- Tên cơ quan tổ chức: viết tắt các cơ quan tổ chức nước ngoài thông dụng

Ví dụ: *UNESCO; UNDP; UNICEF; v.v...*

Nếu tên các cơ quan tổ chức nước ngoài chưa thông dụng thì không được viết tắt mà phải viết tên đầy đủ bằng tiếng Việt.

Ví dụ: Thư viện quốc gia Nga (không viết tắt RSL); Thư viện Quốc hội Mỹ (không viết tắt LC); v.v...

Không viết tắt tên các cơ quan tổ chức Việt Nam.

2.4.2. Yêu cầu đối với chuyên gia định từ khóa

a. Yêu cầu về nghiệp vụ thông tin tư liệu:

Chuyên gia định từ khóa cần có kiến thức và kỹ năng cơ bản tối thiểu về nghiệp vụ thông tin tư liệu, bao gồm:

- kiến thức chung về xử lý thông tin;
- kiến thức chung về tổ chức hệ thống tìm tin tư liệu nói chung và về chức năng nhiệm vụ của hệ thống tìm tin có quan hệ trực tiếp;
- kiến thức và kỹ năng thực hiện công đoạn định từ khóa;
- kiến thức và kỹ năng về tìm tin tư liệu;
- kiến thức và kỹ năng về người dùng tin của hệ thống tìm tin có quan hệ trực tiếp và nhu cầu tin của họ.

b. Yêu cầu về kiến thức ngôn ngữ:

- kiến thức và kỹ năng về xử lý thông tin trong mối quan hệ ngôn ngữ của văn bản nội dung tài liệu;
- kiến thức và kỹ năng về xử lý từ vựng.

c. Yêu cầu về ngoại ngữ:

Xuất phát từ đặc điểm tài liệu khoa học là đa dạng về các ngôn ngữ được sử dụng, chuyên gia định từ khóa cần sử dụng thành thạo (đọc và hiểu) một/một số ngoại ngữ được sử dụng trong các tài liệu được xử lý.

d. Yêu cầu về trình độ chuyên môn phù hợp với lĩnh vực nội dung tài liệu

Chuyên gia định từ khóa cần có trình độ chuyên môn nhất định về lĩnh vực nội dung tài liệu để có thể hiểu chính xác nội dung tài liệu, chọn lựa chính xác và đầy đủ các từ khóa phản ánh nội dung tài liệu.

2. PHƯƠNG PHÁP ĐỊNH TỪ KHÓA

Quy trình định từ khóa bao gồm các công đoạn (các bước) cơ bản sau đây:

- Phân tích nội dung tài liệu;
- Chọn từ khóa mô tả;
- Hoàn chỉnh quá trình định từ khóa.

2.1. Phân tích nội dung tài liệu

a. Mục đích: mục đích của phân tích nội dung tài liệu khoa học để định từ khóa là xác định rõ tài liệu đề cập đến đối tượng gì và đối tượng này được đề cập như

thế nào? Nói cách khác, là xác định rõ đối tượng bậc 1, đối tượng bậc 2, các phương diện nghiên cứu của đối tượng và phương pháp nghiên cứu đặc thù được đề cập trong nội dung tài liệu.

b. Phương pháp phân tích nội dung tài liệu

b.1. Đọc tài liệu

Các yếu tố cần đọc trong một tài liệu khoa học:

- *Nhan đề tài liệu*: nhan đề tài liệu thường cung cấp thông tin về đối tượng bậc 1, bậc 2 và phương diện của đối tượng. Tuy nhiên, nếu nhan đề tài liệu được đặt gián tiếp theo chủ đề hoặc đặt theo cảm xúc thì việc đọc nhan đề không giúp xác định được đối tượng và phương diện của đối tượng. Hiện tượng này thường xảy ra đối với tài liệu khoa học xã hội. Tuy nhiên cũng không phải là ngoại lệ đối với các tài liệu khoa học tự nhiên, khoa học kỹ thuật và công nghệ;

- *Mục lục tài liệu*: việc đọc mục lục tài liệu có thể giúp xác định đối tượng bậc 1, đối tượng bậc 2, các phương diện bậc 1 và bậc 2 của đối tượng. Như vậy, đọc mục lục tài liệu có thể giúp người định từ khóa xác định toàn bộ các yếu tố cơ bản của nội dung tài liệu. Tuy nhiên, không phải tài liệu khoa học nào cũng có mục lục (ví dụ các bài báo của tạp chí khoa học, báo cáo hội nghị, hội thảo, tiêu chuẩn, sáng chế phát minh v.v...thường không có mục lục). Như vậy, việc đọc mục lục để xác định nội dung tài liệu thường chỉ áp dụng đối với tài liệu khoa học dạng sách, chuyên khảo...;

- *Toàn văn nội dung tài liệu*: đây là yếu tố tin cậy nhất giúp xác định các đối tượng và phương diện của

đối tượng. Tuy nhiên, việc đọc toàn văn tài liệu chỉ khả thi đối với các tài liệu có khối lượng nhỏ như các bài báo khoa học, các báo cáo hội nghị, hội thảo... Khi đọc toàn văn nội dung các tài liệu này, các yếu tố cần chú ý là các đề mục (nếu có). Đề mục có thể cung cấp những thông tin về đối tượng bậc 2 và phương diện bậc 2 của đối tượng bậc 1;

- *Các yếu tố khác*: một số yếu tố khác của tài liệu khoa học có thể cung cấp những thông tin tham khảo, giúp chuyên gia định từ khóa trong việc xác định nội dung tài liệu. Các yếu tố này là: bài tóm tắt của tác giả tài liệu; lời giới thiệu; kết luận; các câu, đoạn văn có dấu hiệu đặc biệt (gạch chân, in đậm, in nghiêng...). Tuy nhiên, thông tin thu được từ các yếu tố này thường chỉ có giá trị tham khảo chứ không có giá trị quyết định.

b.2. Xác định các đối tượng, phương diện nghiên cứu và phương pháp nghiên cứu đối tượng

Quy trình xác định đối tượng và phương diện của đối tượng như sau:

- bước 1: xác định đối tượng bậc 1;
- bước 2: xác định đối tượng bậc 2;
- bước 3: xác định các phương diện bậc 1 và bậc 2 của đối tượng và xác định các phương pháp nghiên cứu đặc thù;
- bước 4: liệt kê các đối tượng và phương diện đã được xác định. Đánh dấu các đối tượng bậc 2 có trọng số thông tin cao.

c. Kết quả của giai đoạn phân tích nội dung tài liệu:

- xác định được đối tượng bậc 1 và đối tượng bậc 2;
- xác định được phương diện bậc 1 và phương diện bậc 2;
- xác định được phương pháp nghiên cứu.

2.2. Chọn từ khóa mô tả

a. Mục đích: chuyển đối tượng, phương diện đối tượng và phương pháp nghiên cứu được thể hiện bằng ngôn ngữ tự nhiên của tài liệu sang dạng thể hiện bằng từ khóa.

b. Phương pháp:

b.1. Phương pháp định từ khóa tự do: trên cơ sở các đối tượng, phương diện đối tượng và phương pháp nghiên cứu đã xác định, chuyên gia định từ khóa tự mình lựa chọn các từ khóa để mô tả mà không có sự trợ giúp của các phương tiện kiểm soát từ vựng. Việc định từ khóa tự do có thể thực hiện theo các bước như sau:

* *Bước 1:* xem xét toàn bộ các đối tượng, phương diện đối tượng và phương pháp nghiên cứu đối tượng về mặt từ vựng, căn cứ vào các yêu cầu đối với nội dung và hình thức từ khóa để xử lý từ vựng và lựa chọn từ khóa mô tả. Có thể có 2 trường hợp:

- đối tượng, phương diện đối tượng, phương pháp nghiên cứu đã được tác giả tài liệu mô tả theo đúng các yêu cầu đối với thuật ngữ khoa học, tức là về mặt từ vựng đã đảm bảo các yêu cầu về từ khóa;

- đối tượng, phương diện đối tượng, phương pháp nghiên cứu chưa được mô tả theo đúng các yêu cầu đối với thuật ngữ khoa học. Khi đó, chuyên gia định từ khóa tự mình lựa chọn từ khóa mô tả bằng cách chỉnh sửa hoặc thêm vào các từ mới, căn cứ vào:

+ kiến thức của chuyên gia định từ khóa về vấn đề nội dung tài liệu đề cập; kiến thức về ngôn ngữ; kinh nghiệm định từ khóa v.v...;

+ các yêu cầu về nội dung và hình thức từ khóa;

+ các phương tiện trợ giúp như: từ điển thuật ngữ chuyên ngành, tài liệu hướng dẫn...

* *Bước 2:* liệt kê các từ khóa đã chọn;

* *Bước 3:* phân loại các từ khóa đã chọn thành từ khóa chính và từ khóa phụ theo các tiêu chí đã trình bày ở mục b, c của đề mục 1.3 trong tài liệu này.

b.2. Phương pháp định từ khóa có kiểm soát: trên cơ sở đối tượng, phương diện đối tượng và phương pháp nghiên cứu đã xác định, chuyên gia định từ khóa chọn từ khóa để mô tả chúng với sự trợ giúp của các phương tiện kiểm soát từ vựng. Các phương tiện này có thể là:

- Từ điển từ khóa đa ngành hoặc chuyên ngành;
- Thesaurus đa ngành hoặc chuyên ngành;
- Bảng tra chủ đề (Subject Heading).

Việc định từ khóa có kiểm soát được thực hiện theo các bước như sau:

* *Bước 1:* về cơ bản giống như phương pháp định từ khóa tự do.

* *Bước 2:* sử dụng phương tiện kiểm soát từ vựng để kiểm soát từ khóa như sau:

Chuyên gia định từ khóa đối chiếu các từ ngữ mô tả đối tượng, phương diện đối tượng và phương pháp nghiên cứu với các từ thể hiện khái niệm tương ứng có sẵn trong phương tiện kiểm soát từ vựng. Lúc này có 2 trường hợp xảy ra:

- có từ ngữ tương ứng trong phương tiện kiểm soát từ vựng. Khi đó, từ ngữ tương ứng này được chọn là từ khóa;

- không tìm thấy từ ngữ tương ứng trong phương tiện kiểm soát. Khi đó cần xem xét các nguyên nhân sau:

- + Do đối tượng, phương diện đối tượng và phương pháp nghiên cứu của tài liệu chưa được mô tả đúng với từ ngữ mang khái niệm tương ứng trong phương tiện kiểm soát từ vựng hoặc mô tả không phù hợp với các quy ước về từ vựng của phương tiện kiểm soát từ vựng. Khi đó phải xem xét lại các yêu cầu về từ khóa để mô tả cho phù hợp;

- + Do đối tượng, phương diện đối tượng nghiên cứu của tài liệu là những cái mới chưa được đưa vào phương tiện kinh từ khóa xử lý như trong trường hợp định từ khóa tự do và nên tham khảo thêm ý kiến các chuyên gia trong lĩnh vực đề tài của tài liệu. Đồng thời xem xét để bổ sung kịp thời từ khóa mới vào phương tiện kiểm soát từ vựng.

- + Do đối tượng hoặc phương diện đối tượng nghiên cứu của tài liệu quá sâu so với diện bao quát đề tài của

hệ thống tìm tin cụ thể. Khi đó, cần tìm từ khóa mô tả đối tượng/phương diện đối tượng bậc cao hơn gần nhất trong phương tiện kiểm soát từ vựng. Tuy nhiên, việc xác định khi nào đối tượng/phương diện đối tượng được coi là quá sâu so với diện bao quát đề tài của hệ thống tìm tin cần p^o khối lượng tài liệu lớn, người dùng tin sẽ buộc phải chọn lựa một lần nữa kết quả tìm tin để tìm được những tài liệu đề cập đến những đối tượng mà người định từ khóa coi là quá sâu so với diện bao quát đề tài của hệ thống.

* *Bước 3*: liệt kê từ khóa đã chọn.

* *Bước 4*: phân định từ khóa chính và từ khóa phụ (tương tự như trong phương pháp định từ khóa tự do).

* *Bước 5*: bổ sung từ khóa mới xuất hiện vào phương tiện kiểm soát từ vựng. Việc bổ sung này có thể thực hiện thường kỳ hoặc định kỳ tùy theo số lượng thuật ngữ mới xuất hiện trong khi định từ khóa tài liệu.

2.3. Hoàn chỉnh quá trình định từ khóa

Sau khi xác định được các từ khóa của tài liệu bằng phương pháp định từ khóa tự do hoặc phương pháp định từ khóa có kiểm soát, việc tiếp theo là hoàn chỉnh quá trình định từ khóa. Mục đích của công đoạn này là đảm bảo độ đầy đủ, chính xác và hiệu suất đáp ứng yêu cầu tin của các từ khóa đã chọn. Cụ thể là:

a. Hoàn chỉnh về mặt định lượng

Kiểm tra xem các từ khóa đã phản ánh đầy đủ các yếu tố cần thiết của nội dung tài liệu hay chưa bằng cách đối chiếu các từ khóa đã chọn với kết quả phân tích nội dung tài liệu. Nếu chưa đầy đủ, phải bổ sung các từ khóa còn thiếu;

b. Hoàn chỉnh về mặt định tính

b.1. Trong một số trường hợp khi định từ khóa theo phương pháp tự do có thể xem xét từ vựng mang khái niệm tương đương với từ vựng mô tả đối tượng bậc 1. Sự tương đương này xét theo khía cạnh từ vựng là hiện tượng đồng nghĩa và giả đồng nghĩa của từ. Việc bổ sung này thường chỉ áp dụng trong trường hợp từ khóa bổ sung phải là các thuật ngữ được sử dụng ở mức phổ biến gần như từ khóa chính.

Ví dụ:

+ bổ sung theo khía cạnh đồng nghĩa:

Đối tượng (từ khóa chính): Từ khóa bổ sung :

Lợn

Heo

Quyền con người

Nhân quyền

Thuốc diệt chuột

Bả chuột

Bệnh da liễu

Bệnh ngoài da

Doanh nghiệp quốc doanh

Doanh nghiệp nhà nước

Mạng cục bộ

Mạng LAN

v.v...

+ bổ sung theo khía cạnh giả đồng nghĩa:

Đối tượng (từ khóa chính): Từ khóa bổ sung:

Mối mọt Chống mối mọt

Ô nhiễm Chống ô nhiễm

Mù chữ Xóa mù chữ

Nghèo Giảm nghèo

v.v...

d. Trong một số trường hợp, có thể xem xét bổ sung từ khóa theo quan hệ liên đới của từ vựng mô tả đối tượng bậc 1 (căn cứ vào thực tế nội dung tài liệu)

Ví dụ:

Đối tượng (từ khóa chính): Từ khóa bổ sung:

Sợi bazan Sợi cách nhiệt

Trồng rừng Rừng

Lập kế hoạch Kế hoạch

Công nghiệp thực phẩm Thực phẩm

An toàn phóng xạ Phóng xạ

v.v...

e. Ngoài ra, trong trường hợp đối tượng/phương diện đối tượng nghiên cứu của tài liệu mới xuất hiện, cần xem xét đưa chúng vào nhóm/linh vực phù hợp trong Khung phân loại hoặc Khung đề mục mà Hệ thống thông tin tư liệu đang sử dụng, đồng thời có thể xem xét bổ sung thêm từ khóa

mô tả đối tượng/phương diện đối tượng bậc cao hơn gần nhất trong phương tiện kiểm soát từ vựng vào danh sách từ khóa của tài liệu.

2.4. Trình bày từ khóa trong biểu ghi

Sau khi đã hoàn tất các công đoạn của quy trình định từ khóa tài liệu, các từ khóa nên được trình bày theo trọng số thông tin của chúng trên các Phiếu xử lý tiền máy (Worksheets) và Biểu ghi (Records) của hệ thống tìm tin.

a. Mục đích: tăng cường vai trò thông báo về nội dung tài liệu của các từ khóa, nhằm định hướng cho người dùng tin về những vấn đề cơ bản của nội dung tài liệu. Mặc dù đối với một vài hệ thống tìm tin, thứ tự sắp xếp các từ khóa là không quan trọng. Nhưng đối với người dùng tin, nếu các từ khóa của một tài liệu được sắp xếp trên biểu ghi một cách có chủ ý để định hướng cho họ về những vấn đề quan trọng của tài liệu thì sẽ giúp họ có được những nhận dạng cơ bản ban đầu về nội dung tài liệu mà chưa cần tham khảo các yếu tố khác của biểu ghi.

b. Thứ tự sắp xếp từ khóa trong biểu ghi

Các từ khóa nên sắp xếp theo thứ tự ưu tiên như sau:

- Trường hợp 1: biểu ghi không phân biệt trường từ khóa chính và từ khóa phụ, thứ tự sắp xếp từ khóa như sau:

- + từ khóa chính mô tả đối tượng bậc 1;
- + từ khóa chính mô tả đối tượng bậc 2;

- + từ khóa phụ mô tả phương diện bậc 1;
- + từ khóa phụ mô tả đối tượng bậc 2;
- + các từ khóa phụ còn lại.

- Trường hợp 2: biểu ghi có phân biệt trường từ khóa chính và từ khóa phụ thì các từ khóa được sắp xếp vào các trường tương ứng theo thứ tự ưu tiên như trường hợp 1.

- Trường hợp 3: biểu ghi có các trường từ khóa khác như: từ khóa địa lý, từ khóa nhân vật, từ khóa thời gian...: + Đây là các trường từ khóa phái sinh của trường từ khóa chính và trường từ khóa phụ. Chuyên gia định từ khóa căn cứ vào khái niệm của đối tượng mà từ khóa mô tả để sắp xếp chúng vào các trường tương ứng theo thứ tự như trong trường hợp 1.

3. MỘT SỐ ĐẶC ĐIỂM HỆ THỐNG TÌM TÌN TỪ LIỆU ẢNH HƯỞNG ĐẾN VIỆC ĐỊNH TỪ KHÓA TÀI LIỆU

3.1. Đặc điểm các lĩnh vực nội dung tài liệu khoa học và việc định từ khóa

Bản chất của việc định từ khóa là sử dụng từ vựng khoa học để mô tả nội dung tài liệu, do vậy đặc điểm nội dung tài liệu có ảnh hưởng nhiều, và trong một số lĩnh vực, có thể chi phối việc lựa chọn từ khóa. Mặc dù việc chọn lựa từ khóa chủ yếu là căn cứ vào các yêu cầu chung về nội dung và hình thức từ khóa như đã trình bày ở mục 1.4, phần 1 nhưng việc áp dụng chúng với từng lĩnh vực tài liệu có thể có những điểm khác nhau,

đòi hỏi chuyên gia định từ khóa phải linh hoạt trong từng trường hợp cụ thể. Qua thực tế định từ khóa tại các hệ thống tìm tin tư liệu trong Hệ thống thông tin KHCN quốc gia, có thể đề cập tới một số nhóm tài liệu như sau:

a. Nhóm tài liệu về lĩnh vực khoa học tự nhiên, khoa học kỹ thuật và công nghệ

a.1. Một số đặc điểm cơ bản:

- nội dung thường đề cập đến những đối tượng hữu hình, cụ thể;
- tần số xuất hiện các đối tượng mới thường khá cao so với lĩnh vực khoa học xã hội và nhân văn, đặc biệt là trong một số lĩnh vực khoa học mới và khoa học liên ngành như công nghệ sinh học, công nghệ hóa sinh, y sinh, y dược v.v...;
- nội dung tài liệu thường có đối tượng bậc 2 và phương diện bậc 2 của đối tượng bậc 1.

a.2. Ảnh hưởng của đặc điểm tài liệu tới việc chọn lựa từ khóa

- Ảnh hưởng tích cực :
 - + thuận lợi trong việc xác định đối tượng và phương diện đối tượng khi phân tích nội dung tài liệu.
- Ảnh hưởng tiêu cực:
 - + khó khăn trong việc lựa chọn từ khóa mô tả đối tượng;
 - + khó sử dụng các phương tiện kiểm soát từ vựng trong quá trình định từ khóa.

a.3. Biện pháp khắc phục ảnh hưởng tiêu cực

- cần xây dựng mới và/hoặc thường xuyên cập nhật các thuật ngữ mô tả những đối tượng/phương diện đối tượng mới xuất hiện vào các phương tiện kiểm soát từ vựng;

- thường xuyên tham khảo ý kiến của các chuyên gia về các ngành khoa học liên quan trong khi định từ khóa để xác định mức độ thông dụng của từ khóa trong từng lĩnh vực cụ thể.

b. Nhóm tài liệu khoa học xã hội và nhân văn

b.1. Một số đặc điểm cơ bản:

- đối tượng nghiên cứu của nội dung tài liệu thường là đối tượng vô hình;

- tần số xuất hiện các đối tượng mới thường không cao như trong lĩnh vực khoa học tự nhiên, khoa học kỹ thuật và công nghệ;

- có sự hiện diện của các tài liệu phản ánh hoạt động của các cá nhân, nhóm cá nhân, tổ chức chính trị, tổ chức xã hội... và những tài liệu này, ngoài giá trị khoa học còn mang giá trị lịch sử. Do đó, các yêu cầu về nội dung và hình thức từ khóa, cũng như cách chọn và phân loại từ khóa chính, từ khóa phụ cần được chi tiết hóa thật cụ thể để tránh làm sai lệch nội dung tài liệu và giảm hiệu quả đáp ứng yêu cầu tin của tài liệu.

b.2. Ảnh hưởng của đặc điểm tài liệu tới việc định từ khóa

- Ảnh hưởng tích cực:

+ thuận lợi cho việc sử dụng các phương tiện kiểm soát từ vựng.

- Ảnh hưởng tiêu cực:

+ khó áp dụng các yêu cầu đối với từ khóa.

b.3. Biện pháp khắc phục ảnh hưởng tiêu cực

- cần cân nhắc giá trị lịch sử của một số tài liệu khi áp dụng các yêu cầu đối với từ khóa và phân định từ khóa chính, phụ. Ví dụ: yêu cầu hiện đại, yêu cầu khách quan và yêu cầu thông dụng có thể không hoàn toàn phù hợp với một số tài liệu phản ánh lịch sử hoạt động của Đảng Cộng sản Việt Nam vì trong nội dung tài liệu loại này, có những đối tượng chỉ xuất hiện một lần và gắn liền với ngữ cảnh nội dung tài liệu; Hoặc khi định từ khóa tài liệu phản ánh hoạt động của một tổ chức xã hội nào đó, có thể phải lựa chọn các từ khóa chỉ địa danh cụ thể tới cấp huyện, xã, thậm chí tới tên thôn làng... chứ không phải chỉ hạn chế ở cấp tên nước/tên tỉnh như quy ước thông thường;

- trong trường hợp tài liệu đề cập đến hoạt động của một nhân vật cụ thể nào đó, thì tên nhân vật cần được chọn làm từ khóa chính. Điều này phù hợp với đặc điểm hoạt động nghiên cứu của lĩnh vực khoa học xã hội và nhân văn;

- sử dụng ý kiến chuyên gia, đặc biệt là các chuyên gia am hiểu về giá trị lịch sử của nội dung một số tài liệu đã đề cập.

3.2. Đặc điểm đối tượng người dùng tin và việc định từ khóa tài liệu

Một trong những vai trò quan trọng của từ khóa là đảm bảo cho việc tìm thông tin theo nội dung tài liệu, tức là tìm ra những tài liệu có nội dung phù hợp với yêu cầu của người dùng tin. Do đó, việc hệ thống tìm tin phục vụ người dùng tin nào cũng sẽ chi phối việc định từ khóa. Sự chi phối này xuất phát từ một trong số các đặc điểm yêu cầu tin của những nhóm người dùng tin khác nhau của hệ thống. Đó là *đặc điểm nghề nghiệp*, thể hiện ở chỗ người dùng tin làm những nghề khác nhau thì mục đích tìm tin và độ sâu thông tin của yêu cầu tin sẽ khác nhau. Hay nói cách khác, nghề nghiệp là một trong những yếu tố tạo nên "tập quán tìm tin" của người dùng tin.

Thông thường, người ta phân loại người dùng tin khoa học theo *tiêu chí nghề nghiệp* thành các nhóm chủ yếu như sau:

- người dùng tin là các nhà quản lý;
- người dùng tin là các nhà nghiên cứu, giảng dạy;
- người dùng tin là các nhà sản xuất/kinh doanh, dịch vụ công ích...

Nhóm người dùng tin là các nhà quản lý thường có nhu cầu tin về những vấn đề có tính chất tổng hợp. Do đó, yêu cầu tin của họ có thể không chỉ liên quan đến từ khóa mô tả đối tượng bậc 1 của nội dung tài liệu mà còn đến các từ khóa (hệ khái niệm) mô tả nhóm đối tượng có đặc tính chung. Do đó, nếu đối tượng phục vụ của Hệ thống tìm tin tư liệu chủ yếu là các nhà quản lý thì trong quá trình định từ khóa, vấn đề xem xét bổ sung các từ khóa mô tả đối tượng bậc trên gần nhất của đối tượng bậc 1 là vấn đề cần được chú ý thích đáng;

Nhóm người dùng tin là các nhà nghiên cứu/giảng dạy và những người thuộc khu vực sản xuất, kinh doanh, dịch vụ công ích... thường có nhu cầu tin không chỉ về những vấn đề cụ thể nào đó và các khía cạnh của vấn đề mà còn về những vấn đề khác có liên quan. Do đó, nếu đối tượng phục vụ của Hệ thống thông tin tư liệu chủ yếu là những người thuộc nhóm người dùng tin này thì trong quá trình định từ khóa, cần chú trọng đến những từ khóa mô tả các đối tượng bậc 2 và phương diện của đối tượng bậc 1, đặc biệt khi phân biệt từ khóa chính và từ khóa phụ.

3.3. Đặc điểm tổ chức Hệ thống tìm tin tư liệu và việc định từ khóa tài liệu

Bên cạnh nội dung tài liệu và người dùng tin, một vài đặc điểm tổ chức Hệ thống tìm tin tư liệu cũng có

ảnh hưởng không nhỏ đến việc định từ khóa tài liệu. Đó là:

a. Đặc điểm đối tượng của Hệ thống, đặc trưng bởi loại hình tài liệu mà Hệ thống cập nhật và quản trị. Đặc điểm này thường thể hiện ngay trong tên gọi của Hệ thống tìm tin tư liệu. Đối với mỗi loại hình tài liệu khác nhau, việc định từ khóa cũng có thể khác nhau, nhất là trong công đoạn phân tích nội dung tài liệu, lựa chọn và phân loại, sắp xếp từ khóa. Ví dụ, trong Hệ thống tìm tin tư liệu mà dạng tài liệu là các báo cáo kết quả nghiên cứu khoa học và phát triển công nghệ thì khi chọn từ khóa chính, ngoài đối tượng nghiên cứu, còn phải kể đến phương diện của đối tượng (khía cạnh nghiên cứu), phương pháp nghiên cứu và kết quả nghiên cứu...

b. Đặc điểm tiếp cận của Hệ thống, thể hiện chủ yếu ở phương thức tiếp cận của người dùng tin tới Hệ thống. Nếu người dùng tin tiếp cận trực tiếp (trực tiếp tìm tin qua mạng hoặc trên máy có chứa Hệ thống) thì việc định từ khóa tài liệu khi cập nhật Hệ thống phải tính đến các yếu tố: phương tiện kiểm soát từ vựng trong quá trình định từ khóa và tìm tin; mức độ được đào tạo, huấn luyện về tìm tin của người dùng tin... Trong trường hợp các yếu tố kể trên chưa tới ngưỡng yêu cầu, cần chú trọng tới việc bổ sung các từ khóa theo quan hệ tương đương của thuật ngữ, thậm chí phải

chấp nhận sử dụng một số từ đồng nghĩa khi định từ khóa tài liệu.

4. ĐỊNH TỪ KHÓA YÊU CẦU TIN

Yêu cầu tin là một phần nhu cầu tin của người dùng tin được thể hiện dưới dạng văn bản hoặc lời, là phương tiện thể hiện nhu cầu của người dùng tin về một/một số tài liệu nào đó có nội dung phù hợp với nội dung yêu cầu tin của họ. Việc tìm các tài liệu này có thể được thực hiện bằng nhiều cách, trong đó có cách tìm theo từ khóa. Yêu cầu đầu tiên của việc tìm tin theo từ khóa là phải định từ khóa yêu cầu tin. Việc định từ khóa yêu cầu tin về cơ bản cũng giống như định từ khóa tài liệu, chỉ có điều do khối lượng yêu cầu tin nhỏ hơn nhiều so với tài liệu, nội dung yêu cầu thường khá rõ ràng nên việc định từ khóa vì thế đơn giản hơn nhiều so với định từ khóa nội dung tài liệu. Việc định từ khóa yêu cầu tin có thể được thực hiện theo các bước như sau:

- Bước 1. Cụ thể hóa yêu cầu tin, gồm: *văn bản hóa* yêu cầu tin, nghĩa là ghi ra giấy nếu yêu cầu tin được thể hiện bằng lời nói của người dùng tin và *chính xác hóa* yêu cầu tin nếu yêu cầu tin chưa rõ ràng về nội dung, đặc biệt là về đối tượng của yêu cầu. Có thể vì những lý do chủ quan hoặc khách quan nào đó, người dùng tin chưa thể hiện rõ ràng trong yêu cầu tin là họ muốn tìm những tài liệu có nội dung về vấn đề gì hay cụ thể là về cái gì ? Trong trường hợp này, người phục

vụ thông tin phải bằng mọi cách, xác định cụ thể đối tượng của nội dung yêu cầu tin;

- Bước 2. Phân tích yêu cầu tin: mục đích của bước này là xác định đối tượng của nội dung yêu cầu và các khía cạnh của đối tượng (nếu có). Thực tế, trong nhiều trường hợp đối tượng và khía cạnh của đối tượng đã được làm rõ một phần hoặc toàn bộ ở bước 1. Trong trường hợp như vậy, công việc của bước 2 sẽ dễ dàng rất nhiều và người định từ khóa yêu cầu tin chỉ việc kiểm tra và khẳng định lại đối tượng và khía cạnh của đối tượng yêu cầu tin.

- Bước 3. Lựa chọn từ khóa mô tả đối tượng và khía cạnh của đối tượng

Những công việc của bước này tương tự như trong định từ khóa tài liệu, chuyên gia tìm tin căn cứ vào phương pháp định từ khóa của Hệ thống tìm tin mà thực hiện việc định từ khóa yêu cầu tin. Cụ thể nếu tài liệu của Hệ thống được định từ khóa theo phương pháp định từ khóa tự do hay phương pháp có kiểm soát thì yêu cầu tin cũng được định từ khóa theo phương pháp tương ứng.

Sau khi xác định được các từ khóa của yêu cầu tin, công việc tiếp theo của chuyên gia tìm tin là thiết lập các biểu thức tìm tin bằng các toán tử tìm tin khác nhau nhằm mục đích tìm được đầy đủ và chính xác những tài liệu có nội dung phù hợp với yêu cầu tin.

Trong một số trường hợp, khi định từ khóa yêu cầu tin theo phương pháp có kiểm soát, có thể xuất hiện những thuật ngữ mới. Khi đó, cần nghiên cứu bổ sung những thuật ngữ này vào phương tiện kiểm soát từ vựng.

Cần lưu ý rằng, việc định từ khóa yêu cầu tin không phải nhất thiết lúc nào cũng do các chuyên gia tìm tin thực hiện. Trong trường hợp người dùng tin tự tìm tin trong các cơ sở dữ liệu hoặc qua các mạng thông tin (tìm tin on-line), họ sẽ tự chọn từ khóa tìm tin từ yêu cầu tin của họ. Khi đó, cần có một số đảm bảo cơ bản sau đây để việc tìm tin đạt hiệu quả cao:

- có sự thống nhất về phương pháp chọn từ khóa giữa người dùng tin và người xây dựng hệ thống tìm tin;
- người dùng tin cần được huấn luyện kiến thức cơ bản về hệ thống tìm tin, phương pháp xử lý từ khóa và những kỹ năng nhất định về tìm tin.



TÀI LIỆU THAM KHẢO

1. ISO 1087:1990 - Terminology - Vocabulary.- first edition .-18p.
2. ISO 704:1987 - Principles and methods of terminology .-first edition .-16p.
3. ISO/R 860-1968 - International unification of concepts and terms .-16p.
4. ISO/R 919 -1969 - Guide for the preparation of classified vocabularies (example of method) .-15p.
5. ISO 5963 - Documentation. Methods for examining documents, determining their subjects and selecting indexing terms .-1-st edition, 1985 .-5p.
6. TCVN 5453-1991 - Hoạt động thông tin tư liệu- Thuật ngữ và khái niệm cơ bản.-12tr.
7. TCVN 2243-77 - Chuyển chữ Nga sang chữ Việt .- H.: 1978. Có hiệu lực từ 30/12/1977.
8. Nghiên cứu xây dựng bộ từ khóa cho các cơ sở dữ liệu tư liệu đa ngành khoa học tự nhiên và kỹ thuật/ Báo cáo tổng kết đề tài nghiên cứu khoa học cấp bộ. Mã số 90-60-123.- H.: Trung tâm Thông tin Tư liệu KHCN quốc gia, 1997.-112 tr.
9. Hướng dẫn mô tả nội dung tài liệu bằng từ khóa (định từ khóa).-tài liệu nghiệp vụ/Trung tâm Thông tin Tư liệu KHCN quốc gia, 1993.-27tr.
10. Lê Khả Kế. Về vấn đề thống nhất và chuẩn hóa thuật ngữ trong tiếng Việt .-TC Ngôn ngữ, 1979.-Số 3+4.-tr.27-44;

11. Quy tắc phiên thuật ngữ nước ngoài ra tiếng Việt .-H.: KHXH, 1968 .-16 tr.
12. Galperin I.R. Văn bản với tư cách đối tượng nghiên cứu ngôn ngữ .-Hoàng Lộc d. .-H.: KHXH, 1987.-278tr.
13. Từ điển triết học.- NXB Tiến bộ: Matxcova, 1986 .-Bản dịch.-720 tr.
14. Từ điển tiếng Việt / Văn Tân, Nguyễn Văn Đạm c.l, b.s .- KHXH: Hà Nội, 1994.-912 tr.

MỤC LỤC	Trang
LỜI NÓI ĐẦU	
1. NHỮNG VẤN ĐỀ CHUNG	1
1.1. Một số định nghĩa	1
1.2. Vai trò của từ khóa trong hệ thống tìm tin tư liệu	2
1.3. Phân loại từ khóa	3
1.4. Một số yêu cầu chung	7
1.4.1. Yêu cầu đối với từ khóa	8
1.4.1.1. Yêu cầu về nội dung từ khóa	8
1.4.1.2. Yêu cầu về hình thức từ khóa	16
1.4.2. Yêu cầu đối với chuyên gia định từ khóa	18
2. PHƯƠNG PHÁP ĐỊNH TỪ KHÓA	19
2.1. Phân tích nội dung tài liệu	20
2.2. Chọn từ khóa mô tả	21
2.3. Hoàn chỉnh quá trình định từ khóa	24
2.4. Trình bày từ khóa trong biểu ghi	26
	28

3. MỘT SỐ ĐẶC ĐIỂM HỆ THỐNG TÌM TIN TƯ LIỆU ẢNH HƯỞNG ĐẾN VIỆC ĐỊNH TỪ KHÓA TÀI LIỆU	
3.1. Đặc điểm các lĩnh vực nội dung tài liệu và việc định từ khóa	28
3.2. Đặc điểm đối tượng người dùng tin và việc định từ khóa	30
3.3. Đặc điểm tổ chức hệ thống tìm tin tư liệu và việc định từ khóa tài liệu	32
4. ĐỊNH TỪ KHÓA YÊU CẦU TIN TÀI LIỆU THAM KHẢO	33
	35

