



TRƯỜNG ĐẠI HỌC VĂN HOÁ HÀ NỘI

TRẦN THỊ BÍCH HỒNG - CAO MINH KIỂM

TRA CỨU THÔNG TIN TRONG HOẠT ĐỘNG THƯ VIỆN - THÔNG TIN

Giáo trình dùng cho sinh viên đại học
và cao đẳng ngành Thư viện - Thông tin



NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

TRẦN THỊ BÍCH HỒNG - CAO MINH KIỂM

TRA CỨU THÔNG TIN TRONG HOẠT ĐỘNG THƯ VIỆN - THÔNG TIN

**Giáo trình dùng cho sinh viên đại học
và cao đẳng ngành Thư viện - Thông tin**

(Tái bản có sửa chữa và bổ sung)

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

Mục lục

	<i>Trang</i>
Lời nói đầu	9
Chương 1. Tổng quan về tra cứu thông tin	
1. Tra cứu thông tin	13
1.1. Một số khái niệm	13
1.2. Các dạng tra cứu thông tin	19
2. Hệ thống tra cứu thông tin	29
2.1. Ngôn ngữ tìm tin	30
2.2. Bộ máy tra cứu	70
2.3. Lệnh tìm	110
2.4. Con người	111
Chương 2. Chiến lược tra cứu thông tin	
1. Các bước của quá trình tra cứu	120
1.1. Tìm hiểu yêu cầu tin và phân tích vấn đề	121
1.2. Thể hiện yêu cầu bằng ngôn ngữ tìm tin	125
1.3. Xác định nguồn tra cứu	125
1.4. Lựa chọn công cụ tra cứu	126

1.5. Thực hiện tra cứu	127
1.6. Phân tích kết quả của chiến lược tìm tin	128
1.7. Biên tập và trình bày thông tin	129
1.8. Đánh giá tính phù hợp của thông tin nhận được	130
2. Quá trình tìm tin tự động hoá	137
2.1. Các bước của quá trình tìm tin tự động hoá	137
2.2. Các bước của tìm tin trực tuyến	138
3. Đánh giá hiệu quả tra cứu	139
3.1. Tiêu chuẩn đánh giá	139
3.2. Các chỉ số đánh giá hiệu quả tra cứu thông tin	141

Chương 3. Phương pháp tra cứu thông tin truyền thống

1. Tra cứu thông tin thư mục	144
1.1. Nguồn tra cứu chính	145
1.2. Phương pháp tra cứu	146
2. Tra cứu số liệu và dữ kiện	167
2.1. Nguồn tra cứu chính	167
2.2. Phương pháp tra tìm	168
3. Một số vấn đề cần lưu ý trong quá trình tra cứu	173

Chương 4. Tìm tin tự động hoá

1. Dạng tìm tin tự động hoá	179
2. Cú pháp của tìm tin tự động hoá	181
2.1. Biểu thức tìm	182

2.2. Toán tử	184
2.3. Sử dụng toán tử boole	185
2.4. Toán tử lân cận	192
2.5. Sử dụng ký hiệu chặt cut	198
2.6. Tìm so sánh	199
2.7. Tìm giới hạn theo trường	200
2.8. Trình tự xử lý và thay đổi mức ưu tiên	202
2.9. Sử dụng dấu đóng/mở ngoặc đơn	202
3. Ngôn ngữ lệnh trong tìm tin tự động hoá	203
3.1. Hệ thống dòng lệnh	204
3.2. Hệ thống thực đơn	206
3.3. Hệ thống hỗn hợp	208
3.4. Hệ thống tựa Web	208
4. Những bước tìm tin tự động hoá cơ bản	209
4.1. Xác định các khái niệm và thuật ngữ tìm	211
4.2. Lựa chọn chiến lược tìm tin	213
4.3. Xây dựng biểu thức tìm	214
4.4. Đánh giá sơ bộ và hiệu chỉnh kết quả tìm kiếm	214
5. Chiến lược tìm tin tự động hoá	216
5.1. Chiến lược tìm ngắn gọn	218
5.2. Chiến lược xây dựng khối	218
5.3. Chiến lược các bước liên tiếp	221
5.4. Chiến lược mở rộng dần dần	222

6. Tìm tin trực tuyến	223
6.1. Quá trình phát triển của công nghiệp thông tin trực tuyến	225
6.2. Các thành phần của công nghiệp thông tin trực tuyến	227
7. Tìm tin trên CD-ROM	229
7.1. CD-ROM và cơ sở dữ liệu trên CD-ROM	229
7.2. So sánh tìm tin trên CD-ROM với tìm tin trực tuyến	233
Chương V. Tìm tin trên mạng INTERNET	
1. Những khái niệm cơ bản về Internet	236
1.1. Định nghĩa Internet	236
1.2. Địa chỉ IP và địa chỉ tên miền	238
1.3. Các dịch vụ cơ bản của Internet	242
2. World Wide Web	244
2.1. Một số khái niệm cơ bản của World Wide Web	244
2.2. Trình duyệt Web	249
2.3. Nguồn tin trên World Wide Web	250
3. Tìm tin trên Web	252
3.1. Máy tìm tin	253
3.2. Danh bạ chủ đề	267
3.3. Hướng dẫn chủ đề chuyên môn hoá	269
3.4. Cơ sở dữ liệu	270
Hướng dẫn tự học	272
Tài liệu tham khảo	289

Bảng chữ cái viết tắt

AACR	Quy tắc biên mục Anh - Mỹ (Anglo-American Cataloguing Rule)
BBK	Khung phân loại thư mục thư viện (Biblioteko-Bibliographicheskaja Klassifikatsia)
CD-ROM	Bộ nhớ chỉ đọc trên đĩa nén (Compact Disc Read Only Memory)
DDC	Bảng phân loại thập tiến Dewey (Dewey Decimal Classification)
ISBD	Mô tả sách theo chuẩn quốc tế (International Standard Book Description)
LCC	Bảng phân loại Thư viện Quốc hội Mỹ (Library of Congress Classification)
LCSH	Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ (Library of Congress Subject Headings)
MARC	Biên mục máy tính đọc được (Machine Readable Cataloguing)
MLCĐ	Mục lục chủ đề
MLCC	Mục lục chữ cái

MLPL	Mục lục phân loại
OPAC	Mục lục truy cập công cộng trực tuyến (On-line Public Access Catalog)
RAMEAU	Repertoire d'Autorite' Matiere Encyclope'dique et Alphabetique Unifie
UDC	Bảng phân loại thập tiến quốc tế (Universal Decimal Classification)
UNESCO	Tổ chức Giáo dục, Khoa học và Văn hoá Liên hợp quốc (United Nations's Education, Science and Culture Organisation)
URL	Định danh tài nguyên thống nhất, Định danh nguồn tin thống nhất (Uniform Resource Locator)
WWW	World Wide Web hoặc Web

Lời nói đầu

Từ những năm 50 của thế kỷ XX cho đến nay với sự xuất hiện của nền văn minh trí tuệ thuộc các lĩnh vực khoa học, kỹ thuật, công nghệ, kinh tế và xã hội đã mang lại nhiều biến đổi sâu sắc chưa từng có trong lịch sử loài người. Trên thực tế, thời đại trí tuệ đang được mở màn bởi một loạt các cuộc cách mạng nối tiếp nhau như cách mạng công nghệ, cách mạng thông tin với các ngành có hàm lượng khoa học kỹ thuật cao làm then chốt (tin học, vi điện tử, viễn thông, công nghệ sinh học, vật liệu mới, năng lượng mới,...). Những ngành này đã, đang và sẽ làm biến đổi cơ bản về công cụ, các phương pháp tổ chức quản lý nền sản xuất xã hội cũng như dịch vụ làm cho sản xuất phát triển cao, tinh vi chưa từng thấy. Với sự phát triển của khoa học và kỹ thuật đã dẫn tới sự bùng nổ thông tin. Khối lượng tài liệu khoa học tăng theo cấp số nhân, phong phú về nội dung, đa dạng về hình thức, tồn tại dưới nhiều dạng thức khác nhau (dạng giấy, dạng vi phim, vi phiếu, đĩa từ, đĩa quang...). Đặc biệt sự phát triển cực kỳ nhanh chóng của Internet, đã và đang mở ra những thời cơ và thách thức mới đối với việc sản sinh, lưu giữ, tìm kiếm và cung cấp thông tin.

Chính vì vậy mà việc sử dụng tài liệu và tra cứu thông tin gặp không ít khó khăn, đặc biệt là vấn đề định hướng nguồn tài liệu mới (hàng ngày có khoảng trên 200 trang tạp chí khoa học; mỗi năm có khoảng trên 10 triệu

sáng chế phát minh được công nhận, 5 triệu bài báo được đăng tải và hàng triệu cuốn sách được ra đời). Để giúp người dùng tin định hướng, khai thác và sử dụng một cách triệt để và có hiệu quả các nguồn tin đó, các cơ quan thông tin - thư viện phải có phương pháp tra cứu một cách khoa học và logic.

Tra cứu thông tin là một trong những môn học mang tính khoa học, tính kỹ thuật, tính linh hoạt và định hướng cao, nằm trong chương trình đào tạo ngành Thư viện - Thông tin. Môn học trang bị cho sinh viên những kiến thức cơ bản, có hệ thống về lý luận tra cứu thông tin, cũng như những kỹ năng cần thiết về phương pháp tra cứu. Giáo trình "Tra cứu thông tin trong hoạt động thư viện - thông tin" được biên soạn phù hợp nội dung chương trình giảng dạy đã được thông qua, phù hợp với hoạt động tra cứu trong thực tế rất đa dạng và phong phú đang được tiến hành ở tất cả các thư viện và cơ quan thông tin trong nước, cũng như tiếp cận được những phương pháp tìm tin tự động hoá phù hợp với xu hướng phát triển của ngành trên thế giới. Giáo trình "Tra cứu thông tin trong hoạt động thư viện - thông tin" dùng làm tài liệu giảng dạy và tham khảo cho cán bộ giảng dạy và tài liệu học tập cho sinh viên hệ đại học và cao đẳng ngành Thư viện - Thông tin và trong những trường hợp nhất định có thể làm tài liệu tham khảo cho những người quan tâm tới vấn đề này.

Trong quá trình biên soạn và xuất bản giáo trình chúng tôi nhận được sự đóng góp, giúp đỡ rất chân tình của các đồng nghiệp ở bộ môn Thông tin học, khoa Thông tin - Thư viện Trường Đại học Văn hoá và các chuyên gia trong ngành. Cho phép chúng tôi bày tỏ lời cảm ơn chân thành.

Sau 4 năm xuất bản giáo trình, cùng với sự phát triển mạnh mẽ của khoa học và công nghệ, trong đó có công nghệ thông tin và thông tin học

cũng như thực tế công tác tra cứu trong cơ quan thông tin - thư viện, các tác giả đã cố gắng bổ sung những thông tin mới, cập nhật, đồng thời sửa chữa một số chỗ thông tin chưa chính xác do in ấn. Tuy đã có nhiều cố gắng, chắc chắn giáo trình cũng không tránh khỏi những thiếu sót. Chúng tôi rất mong sự chỉ dẫn và góp ý của đồng nghiệp và độc giả để giáo trình hoàn thiện hơn trong những lần xuất bản sau.

Hà Nội, năm 2008

Các tác giả

Chương 1

Tổng quan về tra cứu thông tin

1. TRA CỨU THÔNG TIN

1.1. Một số khái niệm

Nhiệm vụ chính của mỗi cơ quan thông tin - thư viện là giúp cho người dùng tin/ khách hàng/ bạn đọc định hướng, truy cập, lựa chọn và sử dụng thông tin từ các nguồn tin có trong cơ quan mình và từ các nguồn tin ở nơi khác một cách tốt nhất, thuận tiện và nhanh chóng nhất theo yêu cầu của họ. Ở bất kỳ cơ quan thông tin/ thư viện thuộc hệ thống nào, loại hình nào cũng đều tiến hành giúp bạn đọc định hướng, tra cứu nguồn tìm hoặc hướng dẫn phương pháp giúp họ tự tra cứu thông tin,... Căn cứ vào mục đích và loại hình, các cơ quan thông tin/ thư viện tổ chức phục vụ/ dịch vụ tra cứu một cách khác nhau. Ví dụ: thư viện nhỏ chỉ có 1-2 biên chế, họ phải đảm nhận mọi công việc nghiệp vụ từ khâu bố sung, xử lý kỹ thuật, tổ chức kho tài liệu, tổ chức bộ máy tra cứu, đến phục vụ, trong đó có phần tra cứu; tại thư viện lớn có sự phân công giữa các bộ phận tra cứu - thư mục với các bộ phận

mượn, đọc và giữa các cán bộ trong từng bộ phận với nhau. Công tác tra cứu trong cơ quan thông tin - thư viện bao gồm nhiều khâu công việc như:

- Giúp người dùng tin sử dụng thư viện/ cơ quan thông tin;
- Trả lời các yêu cầu tin;
- Hướng dẫn nghiên cứu các nguồn tin;
- Giới thiệu các sản phẩm và dịch vụ hiện tại;
- Đào tạo người dùng tin,...

Công việc trợ giúp người dùng tin sử dụng thư viện/ cơ quan thông tin.

Nhiều bạn đọc đến cơ quan thông tin, thư viện và cần sự trợ giúp của thủ thư, đặc biệt là thời gian đầu mới đến làm quen và sử dụng thư viện. Có thể họ cần giúp nghiên cứu cách sử dụng hệ thống mục lục, cách thể hiện yêu cầu tin, cách tra tìm tài liệu, nguyên tắc tổ chức, sắp xếp tài liệu trên giá ở các phòng tự chọn, hoặc cách tra cứu trong các tài liệu tra cứu; Có không ít bạn đọc rất quen thuộc với nguồn tin dạng giấy nhưng lại không gần với nguồn tin điện tử. Họ có nhu cầu giúp đỡ định hướng và sử dụng cơ sở dữ liệu, CD-ROM hay Internet. Trong những trường hợp đó thủ thư cần thiết phải giải thích, hướng dẫn họ nghiên cứu, tìm hiểu và cách sử dụng chúng.

Trả lời các yêu cầu tin:

Bạn đọc đến thư viện để tìm thông tin về một tài liệu cụ thể một/ tập hợp tài liệu, số liệu, dữ kiện hay một vấn đề nào đó. Họ cần sự giúp đỡ của thủ thư/ bộ phận tra cứu vì họ không biết bắt đầu xem xét từ đâu? Có những câu hỏi rất đơn giản, ví dụ: Có cuốn sách nào về lịch sử phong trào Cần Vương có trong thư viện không? Những vấn đề

ngiên cứu mang tính tổng hợp như: Những người truyền giáo thế kỷ XVII- XVIII có ảnh hưởng gì đến chính sách phát triển của châu Á - Thái Bình Dương? Cán bộ thư viện cần phải tra cứu thông tin phù hợp để trả lời những câu hỏi đó.

Hướng dẫn nghiên cứu các nguồn tin:

Nếu bạn đọc có câu hỏi mang tính tổng hợp, thủ thư có thể nghiên cứu và hướng dẫn họ sử dụng các nguồn tin khác nhau như: nguồn tin truyền thống có tại thư viện hay những nguồn tin điện tử được khai thác trên cơ sở dữ liệu, các mạng thông tin hoặc internet, không ít trường hợp thư viện/ cơ quan thông tin động viên, giúp đỡ, tạo điều kiện cho họ sử dụng những nguồn tin có thông tin cập nhật và có độ tin cậy cao.

Giới thiệu các sản phẩm và dịch vụ hiện tại:

Các cơ quan thông tin/ thư viện tiến hành nhiều dịch vụ nhằm đảm bảo cho người dùng tin có được thông tin nhanh chóng, phù hợp với nhu cầu và lĩnh vực họ quan tâm, các loại sản phẩm và dịch vụ thông tin gồm:

- Cung cấp tài liệu gốc/ bản sao,
- Danh mục tài liệu/ số liệu/ dữ kiện,
- Biên soạn thư mục, tổng quan về những vấn đề và đề tài phù hợp,
- Phổ biến thông tin chọn lọc,
- Mượn giữa các thư viện,...

Đào tạo người dùng tin

Bất cứ cơ quan thông tin, thư viện nào, dù lớn hay nhỏ đều phải tiến hành đào tạo người dùng tin (tuy nhiên ở các mức độ khác nhau).

Cán bộ thông tin/ thư viện cung cấp những chỉ dẫn, cách sử dụng cơ quan thông tin/ thư viện cho bạn đọc/ người dùng tin cá nhân/tập thể/ nhóm.

Hình thức đào tạo người dùng tin có thể được tiến hành theo các khoá đào tạo khác nhau: theo lớp/ nhóm hoặc tiến hành cho từng đối tượng sử dụng nguồn tin riêng biệt. Đa số cơ quan thông tin/ thư viện trực tiếp hướng dẫn hoặc biên soạn, phân phát tờ rơi giới thiệu các nguồn tin, dịch vụ và sản phẩm của mình cũng như cách khai thác nguồn tin đó - tra cứu thông tin.

Về những vấn đề trên, đã và sẽ được đề cập tới ở nhiều môn học và trong những giáo trình khác nhau trong chương trình đào tạo của ngành. Trong giáo trình này chỉ đi sâu tới vấn đề tra cứu thông tin.

Hiện nay ở Việt Nam thuật ngữ tra cứu thông tin/ tra cứu tin, tìm tin vẫn chưa có sự thống nhất. Trong các cơ quan thông tin thuật ngữ tìm tin được sử dụng tương đối thông dụng, còn thuật ngữ tra cứu được sử dụng và rất quen thuộc trong các thư viện từ lâu. Trong các giáo trình và tài liệu tham khảo của nước ngoài, thuật ngữ tìm tin (information retrieval - tiếng Anh; информационный поиск - tiếng Nga) chủ yếu dùng để chỉ việc tra cứu tự động hoá có sử dụng máy tính điện tử, còn việc tra cứu thông tin theo phương pháp truyền thống, dựa vào các công cụ tra cứu thủ công được gọi là công tác tra cứu (Reference work - tiếng Anh; Справочно - информационная работа - tiếng Nga).

Do quá trình dịch và sử dụng thuật ngữ chưa có sự thống nhất. Ví dụ: Cùng để chỉ hoạt động tra cứu thông tin trong một số tài liệu Anh - Mỹ của các tác giả William A. Katz, Mary Gosling... dùng thuật ngữ

Information retrieval hoặc Charl T. Meadow, Chris Grogan dùng Reference work; Tài liệu tiếng Nga của K. B. Taracanov, A. I. Mikhailop, A. I. Trernui, P. C. Giliarevski,... Информационный поиск hoặc trong các tài liệu tiếng Việt cũng sử dụng các thuật ngữ khác nhau như: PGS. Đoàn Phan Tân, PGS. Phan Văn, TS. Vũ Văn Nhật dùng Tìm tin; các giảng viên chính Trịnh Kim Chi, Dương Bích Hồng và nhiều tác giả khác dùng Công tác tra cứu. Thực sự thuật ngữ *Tra cứu* gần liền và quen thuộc với hoạt động thư viện ở Việt Nam. Tuy nhiên, việc sử dụng thuật ngữ của các tác giả có khác nhau nhưng cùng chỉ một mục đích là quá trình tìm tin/ tra cứu tài liệu/ thông tin nhằm đáp ứng nhu cầu tin của người dùng tin.

Nguyên nhân chính của vấn đề chưa thống nhất thuật ngữ là chúng ta chưa có từ điển chuyên ngành thông tin - thư viện, do đó chưa chuẩn hoá được thuật ngữ.

Trong giáo trình này chúng tôi sử dụng thuật ngữ ***tra cứu thông tin***, song việc sử dụng thuật ngữ trong một số trường hợp không thể thống nhất do ngữ cảnh và bản thân thuật ngữ đó đã được dùng như vậy trong quá trình tìm tin tự động hoá. Ví dụ như: Ngôn ngữ tìm tin.

Khái niệm về tìm tin/ tra cứu thông tin được nhiều tác giả đề cập đến. Sau đây là một số định nghĩa:

- Tra cứu thông tin là tập hợp các công đoạn kỹ thuật và logic với các mục đích cuối cùng là tìm được các tài liệu (bản văn), thông tin về chúng hoặc những sự kiện, dữ kiện riêng biệt về vấn đề mà người dùng tin cần thiết.

- Tìm tin, tra cứu thông tin hay là tập hợp các công đoạn có mục đích, nhằm cung cấp cho người dùng tin những chỉ dẫn hoặc trả lời câu hỏi đột xuất hay thường xuyên của họ.

- Tìm tin là quá trình bao gồm những hoạt động mang tính logic nhằm mục đích cung cấp cho người dùng tin những thông tin phù hợp với yêu cầu của họ theo các dấu hiệu đã có.

- Tìm tin là quá trình so sánh những yếu tố đặc trưng của yêu cầu với những yếu tố đặc trưng của tài liệu nằm trong hệ thống, nhằm xác định sự tương hợp về nội dung, ý nghĩa của các dữ liệu được so sánh và lựa chọn các tài liệu/ thông tin nhằm đáp ứng yêu cầu.

Từ những định nghĩa trên cho thấy các tác giả tương đối thống nhất khi xem xét vấn đề tra cứu/ tìm tin. Như vậy, tra cứu thông tin là một thuật ngữ chung dùng để phản ánh quá trình tra cứu dữ liệu hoặc các nguồn tin, trong đó có cả các số liệu, dữ kiện.

Tra cứu thông tin là quá trình xảy giữa con người và mảng tin thông qua các phương tiện, công cụ/ hình thức lưu trữ thông tin cần thiết khác nhau như hệ thống mục lục, các bảng tra cứu, các ấn phẩm thông tin, các bộ phiếu tra cứu truyền thống/ điện tử, cơ sở dữ liệu... Đó là những công cụ tra cứu thông tin quen thuộc trong các cơ quan thông tin - thư viện. Để tra tìm thông tin cần thiết phải sử dụng một ngôn ngữ tìm tin.

Ví dụ: Ký hiệu phân loại/Chủ đề/Từ khóa.

Hoặc 2, 3 ngôn ngữ tìm tin trên, thông qua các khoá truy nhập/ khoá tra tìm/ điểm tiếp cận thông tin.

✓ Khoá truy nhập/ khoá tra tìm/ điểm tiếp cận thông tin là yếu tố phản ánh những đặc tính khác nhau của đối tượng, được sử dụng cho cả quá trình tra cứu và lựa chọn thông tin. Khoá truy nhập bao gồm:

- ✓ Các thuật ngữ phản ánh đề mục chủ đề, từ khoá (từ chuẩn).

- ✓ Môn loại các đối tượng ngành (chuyên ngành khoa học, loại đơn vị sản xuất kinh doanh, loại đơn vị nghiên cứu, triển khai,...),
- ✓ Thông tin về tác giả,
- ✓ Vật mang tin của tài liệu (tài liệu in, tài liệu điện tử,...),
- ✓ Thông tin về vùng địa lý / địa danh,
- ✓ Nhân vật,
- ✓ Số liệu,
- ✓ Dữ kiện,
- ✓ Ngôn ngữ,
- ✓ Thời gian ...

Có liên quan tới đối tượng mà yêu cầu hướng tới...

1.2. Các dạng tra cứu thông tin

Các dạng tra cứu rất phong phú, chúng được phân chia theo nhiều tiêu chí khác nhau như:

1.2.1. Tính chất thông tin/đối tượng tra cứu

Dựa vào tính chất thông tin của đối tượng tra cứu, có thể phân chia thành các dạng:

- *Tra cứu thông tin thư mục/ Tra cứu thông tin tư liệu:*

Tra cứu thông tin thư mục là quá trình xác định và tách ra khỏi nguồn tìm các tài liệu tương ứng với yêu cầu tin theo các dấu hiệu tìm kiếm cho trước như: tên tác giả, tên tài liệu, nơi xuất bản, nhà xuất bản, năm xuất bản, số trang, số khổ hay bản sao tài liệu gốc. (Tra cứu theo các yếu tố mô tả thư mục).

Trong hoạt động của các cơ quan thông tin thư viện, việc tra cứu thông tin thư mục đã xuất hiện từ rất lâu đời và rất quen thuộc với cán

bộ thư viện và bạn đọc, đó là hình thức cung cấp những thông tin về tài liệu như: tác giả, tiêu đề, lần xuất bản, năm xuất bản, số trang, loại hình tài liệu: sách, báo, tạp chí, bản đồ, bản nhạc, tranh ảnh,... (tài liệu gốc hoặc các bản sao của chúng); thông tin về các sáng chế phát minh/ giải pháp hữu ích/ mẫu mã, nhãn hiệu hàng hóa/ tiêu chuẩn/ mục lục công nghiệp,... cũng có thể một bài trích/ câu trích/ đoạn trích từ sách, báo, tạp chí cụ thể nào đó. Cũng có thể là tài liệu đã qua xử lý (Thông tin cấp 2): Bản mô tả thư mục, bài tổng luận, bản dịch, bài lược thuật,...

Ví dụ:

* Cho biết tên tác giả của tác phẩm: "Hồ Cẩm Đào - Nhà lãnh đạo xuyên thế kỷ của Trung Quốc"?

* Trong thư viện có những tài liệu nào đề cập tới "Mĩ học" xuất bản từ năm 2000 - nay?

* Tài liệu "Thông tin học" của những tác giả nào? Xuất bản vào những năm nào?

* Có thư mục/ Tổng quan nào đã biên soạn về "Vấn đề hội nhập kinh tế của Việt Nam"?

- Tra cứu thông tin dữ kiện:

Tra cứu thông tin dữ kiện là quá trình xác định và tách ra khỏi nguồn tìm các số liệu, dữ kiện tương ứng với yêu cầu tin theo các dấu hiệu tìm kiếm cho trước như tìm các số liệu, dữ kiện có trong bản thân tài liệu (hoặc tập hợp các tài liệu). Hiểu một cách khác, đó là quá trình tra tìm những số liệu, dữ kiện cụ thể như:

- ✓ Đặc tính, tính chất, thông số kĩ thuật của các thiết bị, máy móc,
- ✓ Tính chất của vật, vật liệu,
- ✓ Hằng số vật lý, hoá học,

- ✓ Công thức của một chất/ hợp chất hoá học,
- ✓ Số liệu thống kê,
- ✓ Các khái niệm khoa học,
- ✓ Sự kiện lịch sử, văn hóa,
- ✓ Nhân vật ...

Với sự phát triển không ngừng của nguồn thông tin cấp 1, trong điều kiện các đối tượng cần xử lý thông tin được mở rộng, đặc biệt nhu cầu người dùng tin đòi hỏi cần được cung cấp sản phẩm được tạo nên trên cơ sở xử lý sâu vào nội dung thông tin của nguồn tin, đó là thông tin dữ kiện. Thông tin dữ kiện là các thông tin cụ thể, chi tiết về đối tượng, các thông tin này được thể hiện dưới hình thức các dữ kiện.

Với nền kinh tế mở, nhiều thành phần, các cơ quan thông tin - thư viện, đặc biệt là các cơ quan tư vấn thông tin, trung tâm thông tin hành chính, thông tin thương mại, thông tin kinh tế - xã hội,... rất quan tâm tới hoạt động thông tin dữ kiện. Các cơ quan này đã tạo nên các nguồn thông tin dữ kiện hết sức đa dạng và phong phú, đồng thời cũng là những người dùng tin dữ kiện mạnh nhất. Trong quá trình tra cứu thông tin dữ kiện, dựa vào các yếu tố khác nhau: mục đích sử dụng thông tin, đặc tính của dữ kiện được lưu trữ để đáp ứng yêu cầu của nhiều đối tượng thông qua các công cụ, phương tiện khác nhau như hộp phiếu tra cứu dữ kiện, kho tài liệu tra cứu truyền thống/ điện tử, CSDL dữ kiện,...

Ví dụ:

* Năm 2002 các trường trung học cơ sở ở Việt Nam xây dựng được bao nhiêu phòng học?

* Nhiệt độ lạnh nhất của Sa Pa là bao nhiêu? Những năm nào có tuyết rơi trong 10 năm trở lại đây?

* Có bao nhiêu đại biểu chính thức tham gia Đại hội Đảng lần thứ 3?

* Thủ tướng Thái Lan năm 1997 là ai?

Sự giống nhau và khác nhau giữa tra cứu thông tin thư mục và tra cứu thông tin dữ kiện:

Giống nhau:

- Về nguyên tắc tra cứu.
- Về phương pháp tra cứu.

Khác nhau:

- Về nội dung đối tượng tra cứu.

+ Tra cứu thư mục: Tìm tài liệu hay thông tin về đối tượng/ tài liệu đó.

+ Tra cứu dữ kiện: Tìm thông tin được phản ánh có trong tài liệu/tập hợp tài liệu.

1.2.2. Dựa vào công cụ tra cứu

Công cụ tra cứu/ Nguồn tra tìm/ Bộ máy tra cứu - Đó chính là những nguồn thông tin cấp hai giúp người dùng tin có khả năng tiếp cận tới nguồn thông tin cấp một. Các công cụ tra cứu được lưu trữ bằng các hình thức khác nhau như: hình thức lưu trữ thông tin truyền thống (thủ công), hình thức lưu trữ thông tin bán tự động hoá và hình thức lưu trữ thông tin tự động hoá. Dựa vào các công cụ tra cứu ta có các hình thức cụ thể sau:

- *Tra cứu thông tin truyền thống/ thủ công*: Là quá trình tra cứu được thực hiện thông qua hệ thống tra cứu truyền thống/ thủ công như:

+ Hệ thống mục lục:

- Mục lục chữ cái,
- Mục lục phân loại,
- Mục lục chủ đề, ...

+ Hộp/ bộ phiếu tra cứu:

- Hộp/ bộ phiếu tra cứu cứu chính,
- Hộp/ bộ phiếu tra cứu chuyên đề,
- Hộp/ bộ phiếu tra cứu dữ kiện ...

+ Ấn phẩm thông tin - thư mục:

- Bản thư mục,
- Tạp chí tóm tắt,
- Thư mục trích dẫn tạp chí khoa học,
- Danh mục tài liệu, ...

+ Tài liệu tra cứu:

- Bách khoa thư tổng hợp/ chuyên ngành/ chuyên đề,
- Các loại từ điển,
- Sổ tay tra cứu,
- Niên giám,
- Niên biểu,...

- *Tìm tin bán tự động hoá:*

+ Phiếu lỗ mép.

+ Phiếu-lỗ sợi.

Tuy nhiên, trong thời kỳ công nghệ thông tin và viễn thông phát triển một cách mạnh mẽ và được ứng dụng rộng rãi trong tất cả các lĩnh vực hoạt động, các phương tiện tra cứu bán tự động được tồn tại như một công cụ mang tính lịch sử.

Hiện nay, hầu hết các thư viện và các cơ quan thông tin sử dụng phương tiện tra cứu truyền thống hoặc sử dụng song song với phương tiện tra cứu tự động hoá. Đối với các cơ sở dữ liệu/ ngân hàng dữ liệu tuy mới phát triển trong gần ba, bốn thập kỷ nhưng đã khẳng định được tính ưu việt của mình (mỗi năm tăng trưởng vào khoảng 25%). Tại các thư viện lớn trên thế giới việc tra cứu thông tin được tiến hành chủ yếu bằng phương tiện tự động hoá.

- *Tra cứu thông tin tự động hoá:*

Tra cứu thông tin tự động hoá còn được gọi là tìm tin tự động hoá. Tìm tin tự động hoá là quá trình sử dụng máy tính điện tử và/hoặc mạng máy tính để tìm các thông tin máy tính đọc được, được lưu trữ trên bộ nhớ của máy tính hoặc các thiết bị lưu trữ thông tin điện tử khác và thường được tổ chức dưới hình thức cơ sở dữ liệu.

Tìm tin tự động hoá còn có thể gọi là tìm tin tin học hoá hoặc tìm tin theo chế độ đối thoại bởi đặc trưng rất quan trọng của tìm tin tự động hoá là việc lưu trữ và tìm kiếm thông tin được tiến hành trên cơ sở hệ thống tìm tin có các phương tiện tin học và với sự đối thoại giữa người tìm tin và hệ thống tìm tin. Bộ máy tra cứu trong tìm tin tự động hoá là các cơ sở dữ liệu.

Cơ sở dữ liệu có thể được chia thành:

- Cơ sở dữ liệu thư mục (CSDL TM).
- Cơ sở dữ liệu dữ kiện (CSDL DK).
- Cơ sở dữ liệu toàn văn: Bách khoa thư, từ điển, sổ tay tra cứu, ...

Bản thân tra cứu thông tin/tìm tin tự động hoá có thể được chia thành các dạng khác nhau.

Sự phân loại chi tiết dạng tìm tin tự động hoá được trình bày trong Chương IV: Tìm tin tự động hoá.

1.2.3. Dựa vào hình thức xử lý

Trong quá trình hoạt động, các cơ quan thông tin - thư viện đã sử dụng nhiều hình thức khác nhau để xử lý thông tin như: xử lý hình thức: mô tả thư mục, xử lý nội dung: phân loại tài liệu, định chủ đề, định từ khóa, làm tóm tắt, chú giải, làm tổng luận, tổng quan, lược dịch tài liệu,...

Trên cơ sở các hình thức xử lý có thể chia các loại tra cứu thành:

- *Tra cứu theo dấu hiệu hình thức của tài liệu*: việc tra cứu được thực hiện theo các yếu tố:

+ Tên tác giả/ Dịch giả/ Người chủ biên/ Người sưu tầm/
Người hiệu đính...

+ Tiêu đề tên tài liệu,

+ Theo các yếu tố xuất bản (Nơi xuất bản, nhà xuất bản, năm xuất bản, số trang, số khổ, giá tiền,...).

Ví dụ:

✓ Ai là người dịch cuốn "Chiến quốc sách trong kinh doanh"?
hoặc:

✓ Ai là người sưu tầm và biên soạn cuốn "Trò chơi ảo"?...

- *Tra cứu theo dấu hiệu nội dung*:

Mỗi cơ quan thông tin - thư viện tùy thuộc vào quy mô, điều kiện cơ sở vật chất - kỹ thuật cụ thể hiện có của mình để tiến hành xử lý nội dung tài liệu. Trên cơ sở kết quả của quá trình xử lý nội dung (mô tả nội dung tài liệu): kí hiệu phân loại, đề mục chủ đề, từ khóa/ từ chuẩn,

các cơ quan thông tin - thư viện xây dựng bộ máy tra cứu - công cụ để tiến hành tra cứu. Tra cứu theo nội dung tài liệu/ thông tin có thể được tiến hành:

+ Theo các lĩnh vực tri thức/ môn ngành khoa học :

- Toán học,
- Thiên văn học,
- Điện tử - viễn thông,
- Ngôn ngữ học,...

+ Theo đề mục chủ đề:

- Di tích văn hoá,
- Tiêu hoá,
- Cây công nghiệp,
- Giáo dục,...

+ Theo từ khoá:

- Ăn mòn,
- Tín ngưỡng,
- Máy điện,
- Hạ Long, ...

1.2.4. Dựa vào ngôn ngữ tài liệu

Các thư viện và cơ quan thông tin thường bổ sung tài liệu được xuất bản bằng nhiều ngôn ngữ khác nhau trên thế giới, do đó khi tra cứu cần tìm hiểu người dùng tin cần tài liệu bằng những loại ngôn ngữ nào để đáp ứng yêu cầu.

Vi dụ:

✓ Cho biết tên các tài liệu xuất bản bằng tiếng Lào ở Việt Nam năm 2007?

✓ Có bao nhiêu tài liệu tiếng Nga/ tiếng Nhật/ tiếng Ả-rập/ tiếng Latinh (Anh, Pháp,...) xuất bản ở Việt Nam năm 2004-2006?...

1.2.5. Dựa vào thời gian xuất bản của tài liệu

Trong quá trình tra cứu thông thường người dùng tin cần những thông tin cập nhật qua các nguồn tin mới được xuất bản, phản ánh trình độ phát triển của ngành, lĩnh vực, vấn đề họ quan tâm. Tuy nhiên, không ít trường hợp người dùng tin quan tâm tới những tài liệu được xuất bản trước đó (trong các thời kì khác nhau của lịch sử) hoặc sẽ xuất bản trong thời gian tới. Như vậy, dựa vào thời gian xuất bản của tài liệu ta có thể có các dạng tra cứu:

- Tra cứu thông tin hồi cố (Tài liệu xuất bản trong khoảng thời gian nhất định nào đó đã qua).

Ví dụ:

Các tài liệu về Y học cổ truyền bằng chữ nôm xuất bản trong thời kì phong kiến.

- Tra cứu thông tin hiện tại/ hiện thời (Tài liệu mới đang phát hành hoặc được phát hành trong thời gian ngắn đã qua).

Ví dụ:

Các tài liệu về Nhân quyền ở Việt Nam xuất bản từ năm 2000 đến nay.

- Tra cứu thông tin dự báo (Tài liệu sẽ phát hành trong thời gian tới).

Ví dụ:

Các tài liệu về lịch sử mỹ thuật Việt Nam sẽ xuất bản năm 2009.

1.2.6. Dựa vào loại hình tài liệu

Trong thời kì khoa học - công nghệ phát triển rất mạnh mẽ dẫn tới bùng nổ thông tin, các loại hình tài liệu được xuất bản rất phong phú cả về nội dung cũng như hình thức, đặc biệt là tài liệu khoa học và công nghệ. Ngoài các tài liệu thông thường dạng giấy như sách, báo, tạp chí phản ánh nhiều vấn đề liên quan tới khoa học cũng như đời sống xã hội, còn có những loại hình tài liệu/ thông tin đặc biệt như: Mô tả sáng chế phát minh, mô tả các giải pháp hữu ích, tài liệu tiêu chuẩn - đo lường - chất lượng, mục lục công nghiệp (catalog công nghiệp),... Còn nhiều loại tài liệu dưới dạng vi phim, vi phiếu, tài liệu dưới dạng điện tử cũng ngày càng xuất hiện nhiều hơn và trở nên quen thuộc, thân thiện hơn với người dùng tin. Trong quá trình tra cứu thông tin có thể tra cứu theo loại hình tài liệu riêng biệt.

Ví dụ:

- ✓ Các tiêu chuẩn về xuất khẩu hải sản đông lạnh ban hành năm 2000 - 2003, hoặc:
- ✓ Các văn bằng sáng chế phát minh về Lò điện đã được cấp năm 2000 - 2003.

Trên thực tế, để tra cứu thông tin được nhanh chóng, chính xác và đầy đủ, khi thực hiện quá trình tra cứu không thể chỉ dựa vào một, hai dạng tra cứu riêng biệt, mà chủ yếu là phải phối hợp nhiều dạng tìm với nhau.

Để tra cứu một vấn đề cụ thể, khi sử dụng bộ máy tra cứu cần kết hợp các công cụ tra cứu truyền thống với các công cụ tra cứu tự động hoá; Kết hợp tra cứu các tài liệu như: sách, báo, tạp chí dạng giấy với

tài liệu điện tử; bằng nhiều loại ngôn ngữ khác nhau; thời gian xuất bản tài liệu phù hợp với yêu cầu tin,...

Một trong những yếu tố góp phần đảm bảo cho việc tra cứu thành công và thông tin phù hợp với yêu cầu tin của người dùng tin là người tìm tin phải rất linh hoạt, am hiểu nguồn lực thông tin của cơ quan, có trình độ chuyên môn, nghiệp vụ, có trình độ ngoại ngữ, đặc biệt phải biết sử dụng thông thạo máy tính để có thể khai thác được các cơ sở dữ liệu và nhiều nguồn tin trên các mạng.

Nếu chỉ dựa vào một, hai tiêu chí để tra cứu, kết quả trong nhiều trường hợp sẽ không đầy đủ, không cập nhật; như vậy sẽ không đáp ứng yêu cầu tin của người dùng tin.

2. HỆ THỐNG TRA CỨU THÔNG TIN

Hệ thống tra cứu thông tin là một tập hợp các yếu tố và phương tiện dùng để lưu trữ và tra cứu tài liệu/ thông tin về chúng hoặc những sự kiện, dữ kiện riêng biệt.

Để thực hiện bất kì một dạng tra cứu thông tin nào trong thành phần của hệ thống cũng bao gồm 4 yếu tố cơ bản sau:

- Ngôn ngữ tìm tin,
- Bộ máy tra cứu,
- Yêu cầu tin,
- Con người.

Các yếu tố cơ bản trên của hệ thống có mối quan hệ mật thiết, tác động qua lại lẫn nhau, mật khác có chức năng riêng và tạo nên hệ thống. Chúng ta đi sâu nghiên cứu các yếu tố đó.

2.1. Ngôn ngữ tìm tin

Cơ sở để tạo ra các lệnh tìm đó chính là các loại ngôn ngữ tìm tin. Lệnh tìm là nội dung, ý nghĩa cơ bản của yêu cầu thông tin được diễn đạt bằng những thuật ngữ của ngôn ngữ tìm tin. Nội dung này phù hợp với một tài liệu nào đó hoặc tập hợp nhiều tài liệu khác.

2.1.1. Các khái niệm cơ bản về ngôn ngữ tìm tin

Nhu cầu về sử dụng tài liệu của người dùng tin luôn phát triển cùng với sự phát triển văn hoá, kinh tế, khoa học và kỹ thuật. Nhu cầu khai thác, sử dụng tài liệu/ thông tin không chỉ giới hạn trong phạm vi một vùng, một đất nước thông qua việc trao đổi tài liệu, giao lưu giữa các tổ chức, các cơ quan khoa học, viện nghiên cứu, trường đại học, các thư viện và các cơ quan thông tin trên phạm vi thế giới.

Để người dùng tin khai thác được những thông tin cần thiết, có ích chứa trong tài liệu, các cơ quan thông tin - thư viện phải tổ chức tốt các hoạt động của mình, trong đó có việc xây dựng và áp dụng các ngôn ngữ tìm tin phù hợp.

Do những đặc trưng của ngôn ngữ tự nhiên: tồn tại song song nhiều ngôn ngữ, khác nhau về cấu trúc ngữ pháp, ngữ nghĩa phức tạp, đa dạng nên không thể dùng chúng để xử lý nội dung tài liệu và tra cứu về chúng. Đòi hỏi phải xây dựng ngôn ngữ chuyên dụng để xử lý, tổ chức các hình thức lưu trữ thông tin và tra cứu thông tin. Đó là ngôn ngữ nhân tạo chuyên dụng – *Ngôn ngữ tư liệu/ Ngôn ngữ tìm tin*.

Ngôn ngữ tư liệu: Công cụ dùng để xử lý nội dung tài liệu.

Ngôn ngữ tìm tin: Công cụ dùng để thể hiện yêu cầu tin.

Ngôn ngữ tìm tin là ngôn ngữ nhân tạo do con người quy ước và xây dựng dùng để diễn đạt nội dung, ý nghĩa cơ bản của yêu cầu tin, thể hiện bằng một/ một số tài liệu/ thông tin nào đó đáp ứng yêu cầu tin đặt ra.

Các khái niệm trên được nhiều tài liệu đề cập và đưa ra nhiều định nghĩa với các thuật ngữ khác nhau (Cũng giống như thuật ngữ tìm tin).

Sau đây là một số định nghĩa về ngôn ngữ tìm tin.

- *Ngôn ngữ tìm tin (Retrieval language)*: là phương tiện để ghi lại một cách cô đọng, có khả năng diễn đạt những khía cạnh chủ yếu của yêu cầu tin, được đánh chỉ số hoặc mã hoá theo một hệ thống thích hợp.

- *Ngôn ngữ tìm tin (Retrieval language)*: là ngôn ngữ nhân tạo, các từ của nó là các con số chỉ các khái niệm đã được hệ thống hoá, còn cú pháp phản ánh những quan hệ ngữ nghĩa giữa các khái niệm đó.

- *Ngôn ngữ đánh chỉ số (Index language)*: là ngôn ngữ tập hợp toàn bộ các thuật ngữ đánh chỉ số được sử dụng trong hệ thống tìm tin (các bảng phân loại, bảng đề mục chủ đề, bộ từ khóa/ từ chuẩn).

Tuy nhiên, qua các định nghĩa trên ta thấy có sự chưa thống nhất trong khi sử dụng thuật ngữ. Theo lí giải như ở phần trên, do chưa có từ điển thuật ngữ chuyên ngành. Song qua các định nghĩa có sự thống nhất đó là ngôn ngữ tìm tin là ngôn ngữ nhân tạo chuyên dụng diễn đạt nội dung, ý nghĩa cơ bản của yêu cầu tin và phương tiện để tìm kiếm thông tin. Môn học “Mô tả nội dung tài liệu” chúng ta đã có dịp làm quen với ngôn ngữ tư liệu. Ngôn ngữ tìm tin là thuật ngữ được thay đổi tùy thuộc vào mục đích sử dụng của nó. Tuy nhiên, không ít người chưa hoàn toàn phân biệt được sự giống nhau và khác nhau của ngôn ngữ tìm tin và ngôn ngữ tư liệu.

Ngôn ngữ tìm tin và ngôn ngữ tư liệu giống nhau: Chúng đều là ngôn ngữ nhân tạo chuyên dụng, là công cụ để mô tả nội dung tài liệu và cũng là công cụ để tìm tin.

Ngôn ngữ tìm tin và ngôn ngữ tư liệu khác nhau:

- Ngôn ngữ tư liệu: diễn đạt nội dung, ý nghĩa cơ bản của tài liệu, tạo ra các **mẫu tìm**.

- Ngôn ngữ tìm tin: diễn đạt nội dung, ý nghĩa cơ bản của yêu cầu tin, tạo ra các **lệnh tìm**.

Trong quá trình xử lý nội dung tài liệu tùy thuộc vào mục đích và yêu cầu của thư viện và cơ quan thông tin, cán bộ xử lý tiến hành quá trình phân tích tài liệu và thể hiện bằng những hình thức trình bày của ngôn ngữ tư liệu. Về bản chất, các ngôn ngữ tư liệu được sử dụng nhằm mục đích là nhóm các tài liệu cùng thể hiện một nội dung vào một nhóm chung có cùng dấu hiệu/ ký hiệu, đồng thời tùy theo những đặc điểm cụ thể mà phân nhỏ nội dung của tài liệu ra nhiều góc độ khác nhau. Kết quả của quá trình xử lý nhằm mục đích để tổ chức các hình thức lưu trữ thông tin khác nhau (xây dựng bộ máy tra cứu): tạo các trường để tìm tin trong các cơ sở dữ liệu: trường KHPL, trường từ khóa...; Xây dựng các bảng tra cứu trong các ấn phẩm thông tin: Bảng tra từ khóa (Keyword Index), bảng tra chủ đề (Subject Index), bảng tra tác giả (Author Index)... Các loại ngôn ngữ tư liệu được xây dựng và sử dụng để tạo nên bộ máy tra cứu linh hoạt, có độ tin cậy cao đáp ứng nhu cầu tra tìm thông tin. Ngoài ra, trong một số trường hợp ngôn ngữ tư liệu còn được sử dụng đến khi tổ chức các kho mở và biên soạn thư mục.

Khi cơ quan thông tin - thư viện sử dụng loại ngôn ngữ tư liệu cụ thể nào để xử lý nội dung tài liệu, thì khi thể hiện yêu cầu tin cũng được dùng chính loại ngôn ngữ đó. Như vậy là sự khác nhau của ngôn ngữ tư liệu và ngôn ngữ tìm tin chỉ ở mục đích sử dụng những loại ngôn ngữ đó.

Hiện nay có ba nhóm ngôn ngữ tìm tin chính đang được sử dụng rộng rãi ở các thư viện và cơ quan thông tin là:

- Ngôn ngữ tìm tin theo kí hiệu phân loại (Ngôn ngữ tìm tin theo hệ thống đẳng cấp).
- Ngôn ngữ tìm tin theo đề mục chủ đề.
- Ngôn ngữ tìm tin theo từ khóa/ từ chuẩn.

Ngôn ngữ tìm tin sử dụng để tra cứu thông tin theo nội dung. Có thể thực hiện việc tra cứu theo hình thức truyền thống cũng như hình thức tự động hoá.

Dù có chung một mục đích nhưng mỗi loại ngôn ngữ tìm tin có những ưu thế, đặc điểm riêng. Chúng ta đi sâu nghiên cứu một số loại ngôn ngữ tìm tin chính.

2.1.2. Ngôn ngữ tìm tin theo kí hiệu phân loại (theo hệ thống đẳng cấp)

Ngôn ngữ tìm tin theo kí hiệu phân loại hay còn gọi là ngôn ngữ tìm tin theo môn ngành tri thức/ theo đẳng cấp hệ thống: là loại ngôn ngữ tìm tin cổ nhất và được áp dụng rất rộng rãi (có thể xếp thứ nhất trong các ngôn ngữ tìm tin) ở các thư viện và cơ quan thông tin. Cơ sở để phân loại tài liệu cũng như xây dựng bộ máy tra cứu và thể hiện yêu cầu tin là các bảng phân loại - ngôn ngữ tiền kết hợp.

Việc phân chia các môn ngành tri thức/ lĩnh vực khoa học trong từng bảng phân loại phụ thuộc vào quan điểm và cách phân định của người biên soạn. Theo cấu trúc của các bảng phân loại, trong từng môn loại lớn lại có sự phân chia chi tiết theo đẳng cấp. Quan hệ thứ bậc là loại quan hệ cơ bản đóng vai trò quyết định trong hệ thống phân loại, do đó khi áp dụng việc xây dựng mục lục phân loại, chúng ta sẽ có một hệ thống các dấu hiệu tìm tin theo sơ đồ hình cây.

Hiện nay trên thế giới, tồn tại song song nhiều bảng phân loại khác nhau. Để thuận lợi cho việc xử lý của cán bộ thông tin - thư viện cũng như người dùng tin, cơ quan thông tin - thư viện chỉ sử dụng một bảng phân loại, cũng không hiếm nơi đồng thời sử dụng nhiều bảng phân loại khác nhau, như Trung tâm thông tin Khoa học và Công nghệ Quốc gia sử dụng BBK, Khung Đề mục Quốc gia để phân loại tài liệu và ấn phẩm thông tin, tổ chức vốn tài liệu.

Điểm qua một số bảng phân loại phổ biến nhất hiện nay.

Bảng phân loại thập tiến Dewey (DDC)

Bảng phân loại thập tiến Dewey (Dewey Decimal Classification. Viết tắt là DDC) do nhà thư viện học nổi tiếng người Mỹ Melvin Dewey (1851- 1931) soạn thảo. Bảng phân loại Dewey ra đời do nhu cầu của công tác tổ chức kho sách và công tác thư mục ở Mỹ lúc bấy giờ, nhằm phục vụ kịp thời cho sự phát triển mạnh mẽ của khoa học kỹ thuật và nền sản xuất tư bản ở thế kỷ XIX.

Bảng phân loại Dewey được xuất bản lần đầu vào năm 1876 và được tái bản nhiều lần có bổ sung và mở rộng. DDC được xuất bản thành hai dạng: đầy đủ và rút gọn. Hiện nay, song song với dạng giấy DDC còn được xuất bản dưới hình thức CD gọi là Bảng Dewey điện

từ (Electronic Dewey). Năm 1996, DDC đầy đủ xuất bản lần thứ 21 cùng với “Dewey for Window”. Bảng này bao gồm tất cả các mục đặc trưng của Dewey điện tử nhưng đã được nâng cao cho phù hợp với môi trường Window. DDC rút gọn được xuất bản lần thứ 13. Bảng rút gọn có dung lượng bằng khoảng 2/5 bảng đầy đủ.

DDC được sử dụng rộng rãi nhất trên thế giới (trên 130 nước; dịch sang hơn 30 ngôn ngữ; khoảng trên 200.000 thư viện sử dụng) và là một trong ba bảng phân loại thông dụng của thế giới. DDC có nhiều ưu thế và được cập nhật tương đối kịp thời. Hiện nay DDC được qui định dùng cho hệ thống OCLC. Tại Việt Nam năm 2007, Thư viện Quốc gia đã cho ra đời bản dịch DDC rút gọn (Lần thứ 13), hiện nay các thư viện thuộc hệ thống thư viện công cộng đang triển khai chuyển đổi từ **Bảng phân loại dùng cho thư viện khoa học tổng hợp** sang **DDC rút gọn**.

Cấu trúc của bảng phân loại Dewey

Bảng phân loại Dewey bao gồm bảng chính, 7 bảng trợ ký hiệu và bảng tra chủ đề. Bảng chính DDC gồm 10 lớp cơ bản, ký hiệu bằng 3 chữ số từ 000 – 999 xếp theo thứ tự như sau:

000	Tổng loại
100	Triết học
200	Tôn giáo
300	Các khoa học xã hội
400	Ngôn ngữ học
500	Các khoa học chính xác
600	Các khoa học ứng dụng

700 Nghệ thuật - thể thao

800 Văn học

900 Địa lý. Lịch sử và các khoa học phụ trợ.

Mỗi lớp cơ bản lại chia thành 10 lớp con, mỗi lớp con chia thành nhiều lớp nhỏ và cứ thế tiếp tục, như vậy sự phân chia được tiến hành theo quy tắc thập tiến.

Ví dụ:

500 Các khoa học chính xác

510 Toán học

520 Thiên văn học

.....

590 Động vật học.

Mục 510

511 Toán học

512 Đại số học đại cương

512.1

512.2....

513 Số học

.....

519 Xác xuất và toán học ứng dụng.

Bảng phụ (Bảng trợ ký hiệu)

Trong DDC có 7 bảng trợ ký hiệu với mục đích mở rộng ký hiệu cho các lớp chính.

Các loại bảng phụ của DDC:

1. Bảng các đề mục chuẩn chung (Bảng phụ hình thức).
2. Bảng đề mục địa lý.
3. Bảng phụ văn học (chi phụ cho văn học lớp 8).
4. Bảng phụ ngôn ngữ (cho ngôn ngữ lớp 4).
5. Bảng phụ dân tộc, chủng tộc.
6. Bảng phụ ngôn ngữ: chia nhỏ các nhóm ngôn ngữ như: nhóm ngôn ngữ tượng hình, nhóm ngôn ngữ slavơ... (cho ngôn ngữ lớp 4).
7. Bảng phụ nhân vật.

Các bảng phụ chỉ sử dụng để phối hợp với các bảng chính mà không có giá trị khi đứng độc lập, giúp cho sử dụng bảng phân loại linh hoạt và có hiệu quả hơn.

Bảng phân loại thập tiến quốc tế UDC

Bảng phân loại thập tiến quốc tế UDC (Universal Decimal Classification) ra đời từ năm 1905 do Liên đoàn Tư liệu Quốc tế (FID) biên soạn. UDC được xây dựng trên cơ sở của bảng phân loại DDC, mở rộng (mục kỹ thuật điện tử, kỹ thuật vũ trụ,...) và thay đổi một số môn loại bao gồm tôn giáo, các khoa học xã hội, một số chủ đề chuyên ngành. UDC còn có thêm những quy định mới, nguyên tắc mới về phương pháp liên kết ký hiệu, nhằm tạo điều kiện cho việc phân loại những tài liệu có nội dung phong phú và phức tạp.

Sự khác nhau giữa DDC và UDC thể hiện ở một số khía cạnh nội dung bên trong, còn vẫn giữ nguyên 10 lớp cơ bản từ 0 đến 9 (bỏ các số 0 ở cuối).

Ngoài ra, UDC còn có một hệ thống bảng phụ phong phú để bổ sung cho bảng chính.

Cấu trúc của UDC

Bảng chính:

0. Những vấn đề chung.

1. Triết học.

2. Tôn giáo.

3. Các khoa học xã hội.

4. (Bỏ trống)

5. Khoa học tự nhiên.

6. Khoa học ứng dụng. Kỹ thuật. Y học.

7. Nghệ thuật.

8. Ngôn ngữ. Văn học. Lý luận văn học.

9. Địa lý. Lịch sử. Tiểu sử nhân vật.

Tiếp theo đây cơ bản là những lớp nhỏ hơn được sắp xếp theo một trật tự nhất định, đánh số theo nguyên tắc thập phân. Trật tự các lớp đó phản ánh mối liên hệ giữa chúng theo nguyên tắc từ chung đến riêng.

Ví dụ:

5. Khoa học tự nhiên.

50 Những vấn đề chung về khoa học tự nhiên.

51 Toán học.

52 Thiên văn học.

....

59 Động vật học.

Chi tiết Mục 54: Hoá học.

547 Hoá học hữu cơ.

547.2 Những hợp chất hữu cơ không vòng.

547.2 Hydro cacbon.

547.21 Metan.

Các bảng phụ: Cũng giống như DDC, UDC cũng có các bảng phụ (Bảng trợ ký hiệu). UDC có 7 bảng phụ chia làm hai loại:

Bảng trợ ký hiệu chung: Thể hiện đặc tính hình thức của tài liệu có liên quan tới tất cả các môn ngành khoa học.

Bảng trợ ký hiệu chuyên ngành/ phân tích: phản ánh những đối tượng cụ thể, đặc trưng cho từng ngành.

Các bảng phụ có dấu hiệu nhận dạng riêng và chi sử dụng để phối hợp với các bảng chính mà không có giá trị khi đứng độc lập.

Bảng trợ ký hiệu chung

Bảng phụ hình thức: Phản ánh những khía cạnh hình thức của tài liệu, công dụng của tài liệu như: tài liệu nghiên cứu, sách giáo khoa. Trợ ký hiệu hình thức được viết trong ngoặc đơn với số 0 đứng trước (0...).

(031) Bách khoa thư.

(071) Sách giáo khoa.

- *Bảng phụ địa lý:* Phản ánh những khía cạnh địa điểm địa lý được đề cập đến trong tài liệu, đó là các vùng địa lý tự nhiên các quốc gia, sông, núi, biển, hồ, đại dương... Mỗi khái niệm địa lý có một trợ ký hiệu độc lập, viết trong ngoặc đơn, giống như trợ ký hiệu hình thức nhưng không có số 0 đứng trước.

(4) Châu Âu

(5) Châu Á

Cũng như trợ kí hiệu hình thức, trợ kí hiệu địa lý có thể viết trước ký hiệu chính, trong trường hợp cần sắp xếp tài liệu theo từng vùng địa lý.

- *Bảng phụ dân tộc* giống với trợ kí hiệu ngôn ngữ nhưng để trong ngoặc đơn.

(= 20) Dân tộc Anh

(= 40) Dân tộc Pháp

- *Bảng phụ thời gian*: Phản ánh những khía cạnh về thời gian: thời đại, thế kỷ, năm tháng... kí hiệu thời gian được thể hiện bằng chữ số Ả-rập viết trong ngoặc kép.

Thế kỷ thứ 1 "00"

Thế kỷ thứ 2 "02"

Năm 1975 "1975"

Năm 250 "0250"

Ngày 02 tháng 09 năm 1945: "1945.09.02".

Thời gian trước công nguyên có thêm (-) đứng trước.

Thế kỷ thứ 2 trước công nguyên: "- 2".

- *Bảng phụ ngôn ngữ*: Bảng này dung chữ số Ả-rập từ 0 đến 9 với dấu = đứng trước chữ số:

= 00 Tất cả các ngôn ngữ trên thế giới.

= ...

= 9 Các ngôn ngữ phương Đông, châu Phi và các ngôn ngữ khác.

Bảng trợ ký hiệu chuyên ngành/ phân tích: được sử dụng chủ yếu đối với các tài liệu kỹ thuật, phản ánh những đặc thù riêng đối với loại tài liệu này như: nguyên vật liệu, sản xuất, sửa chữa, ... Bảng này dùng ký hiệu số Ả-rập, bắt đầu bằng dấu chấm và hai số 0.

00 Sản xuất

.002.3 Nguyên liệu

.004.67 Sửa chữa ...

Các dấu hiệu sử dụng trong UDC

Dấu quan hệ (hai chấm:)

Dấu kết hợp (dấu +)

Dấu mở rộng (gạch chéo /) ...

Có các công dụng như trong các bảng phân loại khác.

Bảng phân loại Thư viện - Thư mục (Библиотечно - библиографическая классификация viết tắt BBK).

Bảng phân loại Thư viện - thư mục (BBK) ra đời từ những năm 1960 đáp ứng nhu cầu từ lâu của các thư viện Xô viết, cần có một bảng phân loại phù hợp với nền khoa học kỹ thuật hiện đại, với nền kinh tế và văn hoá xã hội chủ nghĩa.

Nội dung của bảng phải phản ánh được trình độ phát triển của khoa học kỹ thuật hiện đại và tình hình chính trị, xã hội hiện tại. BBK đã mở rộng thêm nhiều lớp cơ bản và phản ánh tương đối đầy đủ các ngành khoa học kỹ thuật mới phát triển trong những năm đầu thế kỷ XX, các ngành đó đều được đưa thành các lớp chính trong bảng phân loại.

Cấu trúc của bảng mềm dẻo, có tính cơ bản và có khả năng phát triển: phản ánh được các ngành khoa học mới; thuận tiện cho việc tra

cứu thông tin truyền thống, cũng như tự động hoá. Bảng phân loại BBK đã được dịch sang tiếng Việt (1975) và được sử dụng ở một số thư viện lớn ở Việt Nam: Thư viện Khoa học và Kỹ thuật Trung ương, Viện Thông tin Khoa học Xã hội, Thư viện Quốc gia Việt Nam và nhiều thư viện khác.

Cấu trúc của BBK

- Các dãy cơ bản

BBK chia tri thức thành 21 môn ngành lớn/ lớp chính thể hiện bằng 28 chữ cái tiếng Nga từ A đến Я, đã chuyển sang chữ cái tiếng Việt A - Z. được phân chia thành 6 nhóm chính.

Nhóm 1: Chủ nghĩa Mác - Lênin.

Nhóm 2: Các khoa học tự nhiên.

Nhóm 3: Các khoa học ứng dụng.

Nhóm 4: Các khoa học xã hội.

Nhóm 5: Các khoa học tư duy.

Nhóm 6: Các vấn đề tổng hợp.

Cụ thể:

A Chủ nghĩa Mác - Lênin.

B Các khoa học tự nhiên nói chung.

C Các khoa học về toán lý.

D Các khoa học về Hoá học.

Đ Các khoa học về Trái đất.

E Các khoa học về Sinh vật.

Ê Các khoa học Kỹ thuật nói chung.

- F Năng lượng. Vô tuyến điện tử.
- G Nghề mỏ.
- H Công nghệ kim loại.
- I Công nghệ hóa học. Sản xuất hóa học. Sản xuất thực phẩm.
- J Công nghệ gỗ. Các ngành sản xuất công nghiệp nhẹ. Ấn loát. Nhiếp, điện ảnh.
- K Xây dựng.
- L Vận tải.
- M Nông - Lâm nghiệp.
- N Y tế. Các khoa học y học.
- Ô Các khoa học xã hội nói chung.
- P Lịch sử. Các khoa học lịch sử.
- Q Kinh tế. Các khoa học kinh tế.
- R Chính trị. Các khoa học chính trị.
- S Nhà nước và pháp quyền. Các khoa học về pháp luật.
- T Khoa học quân sự. Sự nghiệp quân sự.
- U Văn hóa. Khoa học. Giáo dục.
- V Các khoa học ngôn ngữ.
- W Nghệ thuật. Lý luận nghệ thuật.
- X Tôn giáo. Chủ nghĩa vô thần.
- Y Các khoa học triết học. Tâm lý học.
- Z Tài liệu có nội dung tổng hợp.

Tiếp theo đây cơ bản là lớp nhỏ hơn - được xếp theo nguyên tắc từ chung đến riêng, từ lý thuyết đến thực hành. Hệ thống ký hiệu trong BBK dung hỗn hợp cả chữ cái và số. Lớp cơ bản dùng chữ cái in hoa, các lớp tiếp theo là chữ số Ả-rập đánh số theo nguyên tắc thập tiến.

E0 Sinh vật học đại cương.

E1 Cổ sinh vật học.

E4 Vi sinh vật học.

...

E9 Lý sinh học. Hoá sinh học. Sinh lý học động vật và người.

E99 Lý sinh học. Hoá sinh học. Sinh lý học hệ thần kinh và cơ quan cảm giác.

E99.1

Các bảng trợ ký hiệu: Trong BBK các bảng trợ ký hiệu được gọi là các **bảng mẫu**. BBK có 4 bảng mẫu: bảng mẫu chung; bảng mẫu riêng; bảng mẫu sắp xếp; bảng mẫu địa lý và bảng mẫu các dân tộc.

- **Bảng mẫu chung:** phản ánh những khía cạnh phụ của tài liệu được lập đi lập lại ở hầu hết các ngành khoa học: lịch sử các ngành khoa học, phương pháp nghiên cứu, tổ chức lao động, tiêu chuẩn, định mức... Bảng mẫu chung được ký hiệu bằng chữ cái thường kết hợp với chữ số Ả-rập và được ghép với ký hiệu chính bằng dấu ngăn cách chấm (.)

Ví dụ:

D.d: Lịch sử hoá học.

D12d: Lịch sử phát minh các nguyên tố hoá học.

- Bảng mẫu riêng và bảng sắp xếp.

+ *Bảng mẫu riêng*: Dùng để phối hợp với bảng chính của từng lĩnh vực khoa học riêng lẻ, phản ánh những khái niệm được lặp đi lặp lại trong từng lĩnh vực cụ thể. Bảng mẫu riêng của ngành nào chỉ được dùng cho đúng ngành đó. Trong BBK có hơn 130 bảng mẫu riêng ký hiệu bằng chữ số Ả-rập nối liền với khoa học chính bằng dấu gạch ngang (-).

+ *Bảng sắp xếp*: Dùng để sắp xếp các tài liệu theo hình thức xuất bản của các tác phẩm.

Ví dụ:

- ✓ 0 Toàn tập
- ✓ 3 Sách bộ
- ✓ 4 Tuyển tập..

- *Bảng mẫu địa lý và Các dấu dùng trong BBK* Giống như UDC.

- *Bảng mẫu dân tộc*: được sử dụng để phản ánh về các dân tộc trên thế giới. Bảng mẫu dân tộc có ký hiệu là dấu bằng (=) kèm theo một, hai hoặc ba chữ cái đầu tiên trong tên gọi của dân tộc đó để trong ngoặc đơn.

Ví dụ:

- ✓ Các dân tộc Việt Nam (= Vie)
- ✓ Các dân tộc Nga (= Rus)
- ✓ Các dân tộc Thái Lan (= Tha)

Một số bảng phân loại khác

Bảng phân loại của Thư viện Quốc hội Mỹ

Bảng phân loại của Thư viện Quốc hội Mỹ (**Library of Congress Classification** viết tắt là LCC) được xuất bản lần đầu năm 1901 trên cơ sở bảng phân loại Cutter. LCC là tập hợp của các bảng phân loại chuyên ngành độc lập.

Bảng phân loại này sử dụng hệ thống ký hiệu là chữ cái Latinh A đến Z ở lớp cơ bản và lớp 2; Các lớp sau kết hợp chữ cái Latinh với chữ số Ả-rập. LCC có 23 lớp cơ bản. Mỗi lớp trong hệ thống có một cấu trúc riêng.

Cũng như các bảng phân loại đã nghiên cứu ở trên, LCC có hệ thống bảng bổ trợ kí hiệu rất phát triển như: bảng trợ kí hiệu địa lý, ngôn ngữ, dân tộc,... Các bảng trợ ký hiệu gắn liền với từng bảng, từng lớp cụ thể, do đó LCC không có bảng trợ ký hiệu chung.

Bảng phân loại của Thư viện Quốc hội Mỹ là bảng phân loại được biên soạn cho chính Thư viện của mình với vốn tài liệu lớn, nên có cấu trúc chi tiết, tuy nhiên nó không thông dụng và được nhiều thư viện trên thế giới sử dụng. Ở Việt Nam, hiện nay thư viện Trường Đại học Bách Khoa Hà Nội đang sử dụng LCC để phân loại tài liệu thuộc các môn ngành kỹ thuật (Từ 2004).

Trên thế giới vẫn còn tồn tại một số bảng phân loại khác như: **Bảng phân loại Hai chấm của Ranganathan (Colon Classification viết tắt là CC)**, hay ở một số nước dựa vào bảng phân loại UDC, DDC để biên soạn các bảng phân loại cho thư viện của mình.

Ví dụ:

Bảng phân loại dùng cho các thư viện khoa học tổng hợp (Thư viện Quốc gia Việt Nam) xuất bản năm 1969 (Phân thành 17 lớp chính), 1991, 2002 (Phân thành 19 lớp chính). Bảng phân loại này

được sử dụng rất rộng rãi ở các thư viện Việt Nam, đặc biệt hệ thống thư viện công cộng trong nhiều thập kỷ qua.

Khung để mục Quốc gia: do Viện Thông tin Khoa học và Kỹ thuật Trung ương xuất bản năm 1987 trên cơ sở Khung Để mục của Trung tâm Thông tin Khoa học và Kỹ thuật Quốc tế (MockBa). Hiện nay Trung tâm Thông tin Khoa học và Công nghệ Quốc gia đang sử dụng để hệ thống hóa các mảng tin, Phân loại tài liệu và ấn phẩm thông tin.

Bảng phân loại dùng cho thư viện trường phổ thông: Do tác giả Đỗ Hữu Dư biên soạn trên cơ sở Bảng phân loại 19 lớp của thư viện Quốc gia. Bảng phân loại này được dùng cho các thư viện trường học và một số thư viện trường cao đẳng sư phạm trong nước.

Trên thế giới cũng như ở Việt Nam còn sử dụng một số bảng phân loại chuyên ngành, chuyên dạng để phân loại tài liệu và tra cứu thông tin như **Bảng phân loại tiêu chuẩn, Bảng phân loại sáng chế...** Các bảng phân loại đều có cấu trúc tương tự nhau gồm có các bảng chính, các bảng phụ ... Như vậy, để tra cứu thông tin theo môn ngành tri thức một cách chính xác, đầy đủ, ta không chỉ quan tâm tới các *kí hiệu chính*, mà cần phải chú ý tới các *trợ kí hiệu* như: Trợ kí hiệu hình thức,; trợ kí hiệu địa lí; trợ kí hiệu ngôn ngữ; trợ kí hiệu thời gian, công dụng của tài liệu,...

Qua khảo sát các bảng phân loại đang sử dụng, mỗi bảng được xây dựng đều có quan điểm, đặc thù riêng, song các bảng phân loại đều trên đã thể hiện được:

Tính khoa học: Tính chất này thể hiện ở trật tự sắp xếp logic các lĩnh vực tri thức, phản ánh sự phát triển và mối liên quan, cũng như

ranh giới giữa các bộ môn khoa học, phù hợp với phân loại khoa học hiện đại, có hệ thống chi dẫn, chỉ chỗ hoàn hảo, thể hiện ở kết cấu các kí hiệu chặt chẽ và thống nhất, dễ nhớ, dễ sử dụng.

Tính hiện đại: Tính chất này thể hiện trong nội dung các lớp và các lớp con của bảng phân loại, bao hàm được những thành tựu khoa học tiên tiến của thế giới và có khả năng tiên đoán cho tương lai.

Tính phổ cập và mềm dẻo: Đây là khả năng cơ bản và kết hợp của bảng phân loại: nó có thể giản lược cho các thư viện qui mô nhỏ và có thể mở rộng cho các thư viện lớn mà không phá vỡ cấu trúc, có hệ thống các bảng trợ kí hiệu phong phú và bảng tra cứu chủ đề đầy đủ.

Tuy nhiên sử dụng loại ngôn ngữ tìm tin này chúng ta gặp phải một số khó khăn:

- Với việc sắp xếp tài liệu theo hệ thống các ký hiệu phân loại, vấn đề nghiên cứu thường được phản ánh trong nhiều lĩnh vực khác nhau do đó tài liệu về một vấn đề sẽ bị phân tán, xé lẻ ở nhiều môn khoa học khác nhau trong mục lục phân loại.

Ví dụ:

✓ Gạo

6C8.155 - Chế biến bột

613.26 - Thực phẩm học

6C9.85 - Thương phẩm học

6C8.11 - Xay xát

333.278 - Xuất khẩu

✓ Quang học

535 - Quang học

551.52 - Hiện tượng quang học trong khí quyển

506.4 - Quang học ứng dụng

605.85 - Dụng cụ quang học

- Do phải quán triệt nguyên tắc phân loại theo môn ngành tri thức thể hiện mối quan hệ thứ bậc trong trật tự lôgic và các bảng phân loại là ngôn ngữ tìm tin tiến kết hợp, do đó chúng không thể hiện các vấn đề mới một cách kịp thời (phải đợi đến các kì chỉnh lí mới có thể bổ sung). Với kết cấu của hệ thống theo sơ đồ hình cây, mục lục phân loại chi đảm bảo cho việc thể hiện *mối quan hệ nội hàm* trong ngành tri thức mà *không phản ánh được mối quan hệ ngoại vi* giữa các vấn đề.

- Dù cho các bảng phân loại được biên soạn đều dựa trên nhiều cơ sở khoa học nhất định và có những ưu thế khác nhau về phương pháp luận, song không thể tránh khỏi những hạn chế trong việc giới thiệu các vấn đề mới hoặc các vấn đề hẹp nhưng được nhiều người quan tâm.

Tuy nhiên, không ai có thể phủ nhận được ngôn ngữ tìm tin theo môn loại cho phép định hướng và tra cứu thông tin theo hệ thống các môn ngành tri thức nhanh chóng, thuận tiện và đầy đủ nhất.

2.1.3. Ngôn ngữ tìm tin theo đề mục chủ đề

Ngôn ngữ tìm tin theo đề mục chủ đề là ngôn ngữ tìm tin tồn tại song song với ngôn ngữ tìm tin theo môn loại, hiện nay được sử dụng khá phổ biến ở các nước phát triển. Đây là loại ngôn ngữ tìm tin vừa có tính tiến kết hợp vừa có tính hậu kết hợp.

Với việc sử dụng ngôn ngữ tìm tin theo đề mục chủ đề, cho phép người dùng tin có điều kiện tiếp cận với vốn tài liệu/ thông tin theo vấn đề. Trong ngôn ngữ tìm tin theo đề mục chủ đề mối quan hệ quan trọng nhất đó là mối *quan hệ toàn thể - bộ phận*. Với **nguyên tắc chia nhỏ chủ đề và nguyên tắc diện hoá, vấn đề được chi tiết hoá** ra thành nhiều bộ phận và nhiều diện nghiên cứu khác nhau.

Mối quan hệ diện và mối quan hệ toàn thể bộ phận trong ngôn ngữ tìm tin theo chủ đề được thể hiện bằng mối quan hệ giữa đề mục chủ đề và phụ đề (tiểu đề mục).

Với phương pháp tổ chức hình thức lưu trữ theo chủ đề, tài liệu về cùng một vấn đề được tập trung vào một chủ đề với các phụ đề, thể hiện các góc độ nghiên cứu khác nhau của vấn đề, như vậy nó thể hiện ưu thế phản ánh tài liệu tập trung theo vấn đề, khả năng cập nhật các khái niệm mới, cũng như giới thiệu các vấn đề mới.

Ví dụ:

Gạo (chủ đề)

- | | |
|-------------------|---------|
| - Chế biến bột | 6C8.155 |
| - Thực phẩm học | 613.26 |
| - Thương phẩm học | 6C9.85 |
| - Xây xát | 6C8.11 |
| - Xuất khẩu | 333.278 |

Như vậy, những vấn đề có liên quan tới Gạo sẽ được tập trung vào một chỗ, thể hiện một cách đầy đủ và tạo điều kiện cho việc nghiên cứu vấn đề được thuận lợi hơn so với việc tìm tin theo môn loại (Phân tán ở nhiều lĩnh vực khoa học khác nhau).

Với đặc điểm này, sử dụng ngôn ngữ tìm tin theo đề mục chủ đề, người dùng tin có khả năng truy cập và tìm kiếm thông tin về một vấn đề nhanh chóng và thuận tiện.

Với việc xây dựng ngôn ngữ tìm tin trên cơ sở của ngôn ngữ tự nhiên và theo trật tự vần chữ cái là dấu hiệu cơ bản để sắp xếp các đề mục chủ đề trong mục lục chủ đề, cũng như bảng tra chủ đề chữ cái trong các ấn phẩm thông tin, hoặc ô tra chủ đề dùng cho mục lục phân loại, do đó khi có khái niệm mới hoặc vấn đề mới xuất hiện, chúng ta có thể dễ dàng bổ sung và cập nhật vào hệ thống và tra cứu được tài liệu thích hợp về các vấn đề, về khái niệm đó.

Với các ưu thế trên, việc tra cứu thông tin được tiến hành một cách dễ dàng, nhanh chóng và thuận lợi.

Hiện nay các thư viện, cơ quan thông tin trên thế giới đang sử dụng một số bảng đề mục chủ đề để định chủ đề và tra cứu theo chủ đề. Chúng ta khảo sát một số bảng đề mục chủ đề chính:

Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ

Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ (**Library of Congress Subject Headings** viết tắt là **LCSH**) hiện nay đã được xây dựng trên cơ sở bảng đề mục chủ đề mà thư viện đã soạn thảo từ năm 1898 dựa trên nguyên lý của Charler Ammi Cutter (do Ủy ban Hội thư viện Mỹ biên soạn).

Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ xuất bản lần thứ nhất trong khoảng từ năm 1909 - 1914 với tên gọi "**Đề mục chủ đề sử dụng cho mục lục kiểu từ điển của Thư viện Quốc hội Mỹ**".

Năm 1919 - Xuất bản lần 2.

Năm 1975 - Xuất bản lần 8 và đổi tên là **Library of Congress Subject Headings: LCSH**).

Năm 1986 - Xuất bản lần 10.

Năm 1993 - Xuất bản lần 16.

Năm 2002 - Xuất bản lần 25.

Năm 2007 - Xuất bản lần 30, gồm 5 tập, 288.000 mục từ.

Mỗi lần xuất bản có bổ sung thêm nhiều chủ đề mới. Ví dụ, năm 1993 - Xuất bản lần 16 có 199.000 đề mục chủ đề, trong đó có 7.000 chủ đề mới và chi dẫn, tham chiếu; có 2.800 chủ đề có sự thay đổi ít nhiều.

Trong **Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ** các đề mục chủ đề được xây dựng trên cơ sở : tên nhân vật, tên các tổ chức, cơ quan, đoàn thể, tên các đối tượng địa lý, địa danh, tên các sự vật hiện tượng các vấn đề, đề tài nghiên cứu... Về mặt từ loại , đề mục chủ đề có thể là một từ hoặc một cụm từ, chúng có thể là danh từ hoặc một tập hợp từ. Trong LCSH ta có thể tìm các dạng đề mục chủ đề:

- Là danh từ (dạng số nhiều)

Dogs (chó).

Children (trẻ em).

Houses (ngựa).

Schools (trường học).

Animals (động vật).

- Là một cụm danh từ (danh từ và tính từ).

Vietnamese Literature (văn học Việt Nam).

Russian Newspapers (báo Nga).

- Là một cụm từ gồm một danh từ kết hợp với một danh từ khác được sử dụng như một tính từ.

Oil Industry (Công nghiệp dầu mỏ).

- Là một cụm từ bao gồm một danh từ kết hợp với một danh từ khác bởi giới từ:

Church in art (nhà thờ trong nghệ thuật).

Human experimentation in medicine (thực nghiệm trên cơ thể người trong y học).

- Là một cụm từ gồm hai danh từ liên kết với nhau bởi liên từ.

Familia and society (Gia đình và xã hội).

Culture and Education (Văn hoá và giáo dục).

Để giúp cho người định chủ đề và tra tìm tin theo chủ đề được chính xác nội dung của tài liệu theo các khía cạnh, góc độ nghiên cứu khác nhau của chủ đề, trong bảng đề mục chủ đề của thư viện quốc hội Mỹ còn lập ra các phụ đề (Subdivisions):

+ Phụ đề đề tài (Topic subdivision).

+ Phụ đề hình thức (Form subdivision).

+ Phụ đề thời gian (Chronological subdivision).

+ Phụ đề địa lý (Geographic subdivision).

Ngoài ra LCSH còn tồn tại hệ thống chi chỗ, chi dẫn rất phong phú:

+ Dùng (Use).

+ Dùng cho (Use for = UF).

+ Nghĩa rộng hơn (Broader terms = BT).

+ Nghĩa hẹp hơn (Narrower terms = NT).

+ Thuật ngữ liên quan (Related terms = RT).

+ Cùng xem (See also = SA).

LCSH được sử dụng rộng rãi ở nhiều nước đặc biệt là các nước nói tiếng Anh.

Ngoài LCSH hiện nay ở Mĩ còn tồn tại hai bảng đề mục chủ đề khác cũng được sử dụng rộng rãi, đó là:

Bảng đề mục chủ đề Sears (Sears list of subject headings).

Bảng này được dùng cho các thư viện công cộng cỡ vừa và nhỏ.

Năm 1923 - Xuất bản lần thứ 1.

Năm 2002 - Xuất bản lần thứ 17.

Bảng đề mục chủ đề dùng cho tài liệu thiếu nhi (subject headings of children's literature)

Năm 1969 - Xuất bản lần thứ 1.

Bảng này cũng liên tục được tái bản và bổ sung.

Bảng RAMEAU (Repertoire d'Autorite' Matiere Encyclope'dique et Alphabetique Unifie viết tắt là RAMEAU) - Danh mục các đề mục chủ đề được sử dụng có tính chất bách khoa và được sắp xếp theo vần chữ cái.

RAMEAU được xây dựng trên cơ sở Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ. Ban đầu Thư viện Quốc gia Canada vùng Kebéch dịch LCSH sang tiếng Pháp để xử lý tài liệu, sau đó Thư viện Quốc gia Pháp đã sử dụng bảng này có bổ sung và biên soạn lại.

Bảng RAMEAU chủ yếu được xuất bản dưới dạng Microfich.

Bảng RAMEAU được sử dụng làm ngôn ngữ tìm tin ở một số nước nói tiếng Pháp như: Thư viện Quốc gia Pháp, Thư viện Quốc

gia Canada, Thư viện Quốc gia Thụy Sĩ. Giống như LCSH, RAMEAU cũng được xây dựng trên cơ sở: tên nhân vật, tên các cơ quan, tổ chức,... Các loại phụ đề như: Phụ đề nội dung, phụ đề hình thức, phụ đề thời gian, phụ đề địa lý và các loại chi dẫn: Xem, dùng cho, thuật ngữ hẹp hơn, thuật ngữ rộng hơn...

Bảng đề mục chủ đề ở Việt Nam

Khác với các thư viện nhiều nước trên thế giới, ở Việt Nam các thư viện không sử dụng bảng phân loại để mô tả nội dung tài liệu và tra cứu thông tin trước, mà lại sử dụng đề mục chủ đề. Định chủ đề và tra cứu thông tin theo chủ đề được quan tâm từ thời Pháp thuộc ở các thư viện lớn như: Thư viện Trung ương Đông dương, Thư viện Trường Viễn đông Bác cổ, Thư viện Trường Đại học Y - Dược,... Tại các thư viện này đã xây dựng hai loại mục lục: Mục lục Chủ đề và Mục lục Tác giả. Mục lục Chủ đề chỉ được xây dựng cho sách tiếng Pháp.

Sau ngày hoà bình lập lại ở miền Bắc (10/1954), các thư viện vẫn tiếp tục tổ chức mục lục chủ đề với tư cách là ngôn ngữ tìm tin chính, tuy nhiên có sự thay đổi cơ bản đó là nó được phản ánh bằng tiếng Việt - quốc ngữ : ngôn ngữ chính thức của nước Việt Nam độc lập. Tuy nhiên, ngôn ngữ tìm tin này chỉ tồn tại không lâu, sau 1960 là ngừng, thay thế cho nó là ngôn ngữ tìm tin theo môn loại. Thời gian gần đây Thư viện Quốc gia Việt Nam cũng đã có chủ trương khôi phục lại việc định chủ đề và xây dựng mục lục chủ đề giúp cho việc tra tìm tin theo nội dung một cách đa dạng hơn và thuận lợi hơn. Như vậy để làm được điều này đòi hỏi phải đầu tư kinh phí, cũng như sức lực rất lớn để xây dựng bảng đề mục chủ đề phù hợp.

Ở Việt Nam một số thư viện đã xây dựng riêng cho mình bảng Đề mục chủ đề để phục vụ cho việc xây dựng bộ máy tra cứu và tìm tin như thư viện Trường Đại học Y Hà Nội.

Bảng Đề mục chủ đề của thư viện Trường Đại học Y Hà Nội gồm hai phần:

Phần 1: Chủ đề chính.

Phần 2: Gồm các trợ ký hiệu ghép với chủ đề chính nhằm chi tiết hoá các khía cạnh, góc độ nghiên cứu của chủ đề chính.

Trong phần 1: Các chủ đề được sắp xếp theo thứ tự vần chữ cái có kèm theo tên của các chủ đề và ký hiệu của chủ đề đó.

Ví dụ:

Bệnh học (và sinh lý bệnh): B1

Chấn thương học: C1

.....

Tiêu hoá: T8

Tra cứu từ điển: T13

Các chủ đề trên có thể được chi tiết bằng các mục nhỏ.

00 - Những vấn đề chung

10 - Mô tả các biểu hiện chung tại các cơ quan bộ phận và nguồn gốc phát sinh,...

Từng mục nhỏ chi tiết hoá tiếp:

00 - Những vấn đề chung

01 - Giáo trình

Thực chất, Bảng đề mục của Thư viện Trường Đại học Y Hà Nội là bảng đề mục chủ đề vẫn còn rất sơ sài, nhiều đề mục vẫn chưa được phản ánh, hoặc phản ánh chưa chi tiết, do đó khi tra cứu chưa thật thuận tiện và đầy đủ. Tuy nhiên, đó cũng là một sự cố gắng lớn của thư viện, phần nào đáp ứng với nhu cầu tìm tài liệu/ thông tin mang tính đặc thù về y học của người dùng tin.

2.1.4. Ngôn ngữ tìm tin theo từ khoá, từ chuẩn

Cùng với sự phát triển mạnh mẽ của công nghệ thông tin, hiện nay các thư viện và cơ quan thông tin trên thế giới cũng đã ứng dụng rộng rãi tin học vào hoạt động của mình. Nhiều ngân hàng dữ liệu, mạng thông tin, hàng vạn cơ sở dữ liệu được xây dựng với nhiều phần mềm tư liệu đã được ra đời giúp cho việc lưu trữ và tra cứu thông tin được nhanh chóng và đầy đủ. Ngôn ngữ tìm tin theo từ khoá, từ chuẩn được sử dụng rất rộng rãi từ những năm 1970 trở lại đây. Đặc biệt ở Việt Nam, sau những năm 1990 hầu hết các thư viện, cơ quan thông tin đều đã ứng dụng công nghệ thông tin vào các khâu hoạt động của mình như: xây dựng cơ sở dữ liệu, in phiếu mục lục, biên soạn thư mục, quản lý bạn đọc, tra cứu thông tin tự động hoá,... Song song với các loại ngôn ngữ tìm tin khác theo nội dung như: chi số phân loại và chủ đề, từ khoá, từ chuẩn được sử dụng như một loại ngôn ngữ tìm tin chính trong các thư viện, cơ quan thông tin, giúp cho việc tìm tin nhanh chóng, thuận lợi và đầy đủ.

Ngôn ngữ tìm tin bằng từ khoá, từ chuẩn giúp tra cứu thông tin có nội dung sâu sắc hơn, cụ thể hơn, chính xác hơn, đầy đủ hơn, với số lượng vô cùng. Để định từ khoá tài liệu và tra cứu thông tin một cách chính xác cần phải xây dựng bộ từ khoá chuẩn/ Từ điển từ chuẩn.

Hiện nay, nhiều cơ quan thông tin - thư viện ở Việt Nam đã xây dựng được các cơ sở dữ liệu (CSDL) phản ánh nguồn tài liệu có tại cơ quan mình hoặc nối mạng thông tin với các thư viện, cơ quan thông tin trong nước, bước đầu chúng ta cũng đã xây dựng được một số các bộ từ khoá, từ điển để tham khảo và sử dụng :

Từ điển từ khoá khoa học và công nghệ

Ở Việt Nam *Trung tâm Thông tin Khoa học và Công nghệ Quốc gia* đã biên soạn **Từ điển từ khoá khoa học và công nghệ**.

Từ cuối những năm 1980 (1986) Trung tâm Thông tin Khoa học và Công nghệ Quốc gia đã tiến hành tin học hoá hoạt động thông tin tư liệu. Trung tâm đã xây dựng các cơ sở dữ liệu (CSDL): CSDL BOOK, CSDL SCITEC, CSDL kết quả nghiên cứu, CSDL STAND, CSDL DETAI, CSDL DINFO, ... và đã tiến hành sử dụng từ khoá để xử lý nội dung tài liệu và làm ngôn ngữ tìm tin phục vụ cho việc tra cứu thông tin tự động hoá.

Thời gian đầu, Trung tâm sử dụng ngôn ngữ từ khoá tự do để xử lý tài liệu, tổ chức lưu trữ và tìm kiếm thông tin. Từ khoá của tài liệu được định tự do từ văn bản hoặc do cán bộ định từ khoá tự đặt ra, ít được tiến hành thẩm tra qua ngôn ngữ có kiểm soát như: các từ điển ngôn ngữ, từ điển thuật ngữ, tiêu chuẩn của Việt Nam về thuật ngữ khoa học, dẫn tới hiệu quả tra cứu tin không cao, do mức độ hình thức hoá thuật ngữ thấp, tính hệ thống hoá không đảm bảo, thông tin bị tán mạn và mang nặng tính chủ quan của người xử lý tài liệu và tra cứu.

Nhằm khắc phục nhược điểm trên, để tăng hiệu quả sử dụng và thống nhất của ngôn ngữ tìm tin bằng từ khoá trong các thư viện và cơ

quan thông tin, năm 1997 Trung tâm Thông tin Khoa học và Công nghệ Quốc gia đã hoàn thành bộ từ khoá đa ngành khoa học tự nhiên và công nghệ, năm 2001 đã xuất bản **Từ điển từ khoá khoa học và công nghệ** (mà tiền thân là một công trình nghiên cứu cấp bộ của Trung tâm được tiến hành từ năm 1993-1997).

Từ điển gồm hai tập (50.000 đơn vị từ vựng).

T.I. Bảng từ khoá theo vần chữ cái.

T.II. Bảng từ khoá hoán vị.

Tập I: Bảng từ khoá theo vần chữ cái.

Trong bảng này, các từ khoá được sắp xếp thành hai nhóm:

- *Nhóm từ khoá chính*: Gồm các từ mô tả chủ đề chính của tài liệu đa ngành khoa học. Theo vần chữ cái tiếng Việt từ A đến Z, theo trình tự phát âm và hệ thống dấu trong tiếng Việt.

Những quan hệ tương đương quy ước giữa các từ khoá cũng được trình bày trong phần này, trong đó các từ không ưu tiên được chỉ dẫn đến từ ưu tiên bằng ký hiệu chỉ dẫn sử dụng. Còn các từ ưu tiên được trình bày quan hệ với các từ không ưu tiên bằng một ký hiệu chỉ dẫn dùng cho.

Ví dụ:

Anten chào

DC Anten parabol

Anten parabol

SD Anten chào

Bệnh xoắn lá

DC Bệnh cuốn lá

DC Bệnh quãn lá

DC Bệnh xuãn lá

Bệnh xuãn lá

SD Bệnh xuãn lá

- *Nhóm từ khoá bổ trợ:*

Gồm những từ có cách viết đặc biệt, được nhóm lại thành từng bảng để tiện tra cứu. Các từ trong bảng cũng được sắp xếp theo thứ tự vần chữ cái và trình tự dấu phát âm như ở phần nhóm từ khoá chính. Các nhóm từ khoá trong bộ từ khoá gồm:

+ Bảng tra địa danh: Là danh mục các từ khoá chi tên các châu lục, các nước, các quần đảo, sông, núi lớn và các khu vực địa lý. Cách phiên âm khác nhau được chỉ dẫn đến cách phiên âm thống nhất theo ký hiệu sử dụng.

Ví dụ:

Hắc Hải

SD Biển đen

Hàn Quốc

DC Nam Triều Tiên

Nam Triều Tiên

SD Hàn Quốc

+ Bảng tra tên các hiệp hội, cơ quan tổ chức nghiên cứu và phát triển công nghệ kinh tế, môi trường ở cấp quốc tế, khu vực và quốc gia lớn gồm: bảng tra tên viết tắt, viết tên đầy đủ.

Ví dụ:

BIEF: Ngân hàng Thông tin Quốc tế về các nước nói tiếng Pháp.

BN: Thư viện Quốc gia (Pháp).

UNESCO: Tổ chức văn hoá - Khoa học giáo dục của Liên hợp quốc.

+ Bảng tra tên đầy đủ - Tên viết tắt.

Ví dụ:

AAEE: Hội kỹ sư điện tử Mỹ.

EUROCEAN: Hội hải dương học châu Âu.

+ Bảng tên thuốc: gồm tên gốc các thuốc chữa bệnh đã được chính thức đăng ký với Tổ chức Y tế thế giới:

Ví dụ: + Ampicilin

+ Atropin

+ Bảng tên sinh vật: gồm tên các sinh vật có giá trị hoặc có hại cần phòng trừ, có mặt tại Việt Nam hoặc có liên quan đến Việt Nam.

Bảng tra tên sinh vật Việt - Latinh.

Ví dụ: *Anas poecilorhynchos*: Vịt trời.

Anemonia: Hải quỳ hoa cúc.

Tập 2: Bảng tra hoán vị.

Bảng tra hoán vị của Từ điển từ khoá khoa học và công nghệ giúp người dùng tin dễ dàng tra cứu được tất cả các từ khoá - là từ cùng chứa một từ được quan tâm, đồng thời cũng có tác dụng làm nổi bật tất cả các từ có ý nghĩa trong từng từ khoá.

Ví dụ:

Khi có yêu cầu tin về "*Chẩn đoán*" ta có thể tìm được từ khoá có liên quan tới "*Chẩn đoán*" như:

Bộ đo *Chẩn đoán*

Dụng cụ *Chẩn đoán*

Máy *Chẩn đoán*

Thiết bị *Chẩn đoán*

Chẩn đoán học

Chẩn đoán huyết thanh

Chẩn đoán kỹ thuật

Chẩn đoán sớm

Chẩn đoán từ xa, ...

Việc hoán vị các từ khoá với một số từ phức tạp dễ bị nhầm lẫn như tên sinh vật và tên loại, họ, bộ, ... sinh vật sẽ giúp cho cán bộ xử lý thông tin và người dùng tin có thể xác định được cách dùng từ chính xác nhất.

Bảng tra hoán vị được trình bày theo các *phần tử được hoán vị*, các phần tử đó được sắp xếp theo vần chữ cái.

Trong bảng tra hoán vị các từ khoa học được hoán vị là:

- Các từ khoá Khoa học - Công nghệ.
- Tên sinh vật bằng tiếng Việt.

Các loại từ khoá không được hoán vị:

- Từ khoá địa lý.
- Tên các hiệp hội tổ chức.
- Tên thuốc.
- Tên các sinh vật bằng tiếng Latinh.

Bộ từ khoá của Thư viện Quốc gia Việt Nam

Bộ từ khoá của Thư viện Quốc gia Việt Nam được biên soạn và xuất bản lần đầu năm 1997 có khoảng 10.000 đơn vị từ vựng; Năm 2005 có bổ sung và xuất bản với số lượng khoảng 20.000 đơn vị từ vựng.

Diện bao quát đề tài của bộ từ khoá mang tính chất đa ngành: liên quan đến tất cả mọi lĩnh vực khoa học: chính trị, kinh tế, văn hoá, khoa học kỹ thuật, văn học nghệ thuật, y học,... gồm các sự vật, hiện tượng, tên các nhân vật, đoàn thể, tổ chức, các địa danh, các vùng địa lý,...

Bộ từ khoá của Thư viện Quốc gia Việt Nam được trình bày trong một tập và được chia ra các bảng:

- *Bảng tra từ khoá nội dung*: Gồm những từ chi tiết, khái niệm thuộc các ngành, các lĩnh vực khác nhau. Từ khoá được trình bày theo trật tự chữ cái và theo trình tự dấu phát âm. Các từ khoá được chia làm hai loại:

+ *Từ ưu tiên*: Là những từ được sử dụng khi định từ khoá hoặc khi tra cứu thông tin.

+ *Từ không ưu tiên*: Là những từ nhằm hướng dẫn người định từ khoá hay người tra cứu hiểu rõ mối quan hệ giữa chúng với từ không ưu tiên.

Ví dụ:

- Danh lam.

Xem: Danh lam thắng cảnh

- Danh lam thắng cảnh

CX: Di tích lịch sử

- Di tích văn hoá

DC: Danh lam

- Danh lam thắng cảnh

+ *Bảng tra từ khoá nhân vật:*

Gồm các nhân vật của Việt Nam và của nước ngoài, được sắp xếp theo vần trật tự chữ cái tên đối với các nhân vật nước ngoài và họ đối với nhân vật Việt Nam, có kèm theo phần năm sinh năm mất. Tên một số nhân vật được thể hiện dưới nhiều bút danh hoặc viết khác nhau do phiên âm được quy vào một tên thường dùng có kèm theo chỉ dẫn "xem" và "DC" (dùng cho).

Ví dụ:

- Bush, G. (1924 -)

Xem: Busơ, G

- Busơ, G. (1924 -)

DC: Bush, G.

- Bethoven, L. (1770 -1827)

Xem: Bêthôven, L.

- Bêthôven, L. (1770 -1827)

DC: Bê - thô - ven, L.

- Beethoven, L.

- Bethoven, L.

+ *Bảng tra từ khoá Địa lý Việt Nam:* gồm hai phần.

Phần 1: Từ khoá sắp xếp theo vần chữ cái các huyện, tỉnh (trong tỉnh có tên các huyện).

Ví dụ:

Hà Nam

Bình Lục

Duy Tiên

Kim Bảng

Lý Nhân

Phù Lý

Thanh Liêm.

Phần 2: Từ khoá sắp xếp theo huyện, mỗi huyện có chú thích thuộc tỉnh nào.

Ví dụ:

Đông Anh (Hà Nội)

Đông Hà (Quảng Trị)

Đông Hưng (Thái Bình)

Đông Sơn (Thanh Hoá)

Đông Triều (Quảng Ninh)

+ *Bảng tra từ khoá địa lý thế giới:* Gồm các từ khoá sắp xếp theo Châu lục: châu Á, châu Âu, châu Đại Dương - châu Úc, châu Mỹ, Châu Phi. Tên địa danh nước ngoài được ưu tiên thông dụng.

Ví dụ:

- Australia

Xem: Ôtxtraylia

- Arhentina

Xem: Achentina

Bộ từ khoá về kinh tế - xã hội

Bộ từ khoá về kinh tế xã hội của Trung tâm Thông tin (Bộ kế hoạch - Đầu tư) được xây dựng và hoàn thành năm 1996, có khoảng 5000 đơn vị từ vựng.

Bộ từ khoá gồm hai tập: bao quát đề tài thuộc hai lĩnh vực:

- *Lĩnh vực kinh tế*: chính sách kinh tế.

+ Điều kiện kinh tế.

+ Các ngành kinh tế: Công nghiệp, Nông nghiệp, Thương mại, Giao thông, Tài chính,...

- *Lĩnh vực văn hoá xã hội*: Văn hoá, Giáo dục, Đào tạo, Lao động, Dân số, Môi trường,...

Tập 1: Từ khoá được sắp xếp theo vần chữ cái và các dấu phát âm (= giống Từ điển từ khóa khoa học - công nghệ). Đi kèm theo từ khoá bằng tiếng Việt là từ khoá bằng tiếng Anh và các ký hiệu phân loại dùng để xác định từ khoá thuộc nhóm nào trong tập 2.

Để thể hiện mối liên hệ với nhau giữa các từ khoá, có thể thể hiện các mối quan hệ tương đương, liên đới và thứ bậc với các chỉ dẫn (giống như Từ điển từ chuẩn của UNESCO):

BT: Nghĩa rộng hơn

NT: Nghĩa hẹp hơn

RT: Nghĩa liên quan

UF: Sử dụng cho

SN: Ghi chú phạm vi

Tập 2: Bảng phân loại tài liệu theo chủ đề và các nhóm từ mô tả:

Trong tập này các từ khoá được sắp xếp theo từng đề mục phân loại. Trong mỗi đề mục, các từ được sắp xếp theo vần chữ cái và vần phát âm.

Bảng phân loại gồm 19 đề mục - được đánh số thứ tự từ 01 - 19:

Ví dụ:

01: Hợp tác quốc tế, quan hệ quốc tế

International cooperation. International relations.

02: Chính sách kinh tế, chính sách xã hội, kế hoạch hoá.

Economic policy, social policy, planing.

.....

.....

19: Thông tin - Tư liệu.

Information, documentation.

Trong mỗi đề mục lại được chia nhỏ thành các tiểu mục.

Ví dụ:

07. Nông nghiệp

07.01. Kinh tế học Nông nghiệp

7.03.01. Các đốn điền

Như vậy, hiện tại ở Việt Nam đang tồn tại và được sử dụng tương đối rộng rãi ba bộ từ khoá trên, ngoài ra cũng tồn tại một số bộ từ khóa khác như: Bộ từ khóa khoa học Xã hội – Nhân văn của Viện Thông tin Khoa học Xã hội; Bộ từ khóa của Trung tâm Thông tin – Thư viện Trường Đại học Y Hà Nội. Bộ từ khóa của Thư viện Sau đại học Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP. Hồ

Chí Minh... tuy nhiên các bộ từ khoá này có số lượng từ vựng không nhiều (Từ điển từ khoá khoa học - công nghệ 50.000 từ; Bộ từ khoá của Thư viện Quốc gia Việt Nam 20.000 từ và Bộ từ khoá về kinh tế xã hội 5.000 từ), không đủ để phản ánh các khái niệm trong các ngành khoa học, các lĩnh vực hoạt động trong giai đoạn hiện nay, chưa nói đến sự phát triển nhanh chóng của khoa học - kỹ thuật và sự ra đời nhiều ngành khoa học mới. Do đó, đa số thư viện và cơ quan thông tin vẫn tiến hành định từ khóa tự do. Do ngôn ngữ tiếng Việt rất đa dạng, phong phú về ngữ nghĩa, cú pháp... vì vậy nếu mỗi thư viện ở các vùng miền khác nhau định từ khoá theo cách của mình, sẽ dẫn tới mất tin trong quá trình tra cứu và ảnh hưởng trực tiếp tới khai thác các mạng thông tin Quốc gia.

Việc xây dựng bộ từ khóa, từ chuẩn, tiến tới xây dựng từ điển từ chuẩn ở Việt Nam tuy đã được nhắc tới trong vài kỳ hội nghị, hội thảo về công tác thông tin - thư viện, song đến nay vẫn chưa được tiến hành triển khai trong phạm vi quốc gia.

Từ điển từ chuẩn của UNESCO

Từ điển từ chuẩn của UNESCO (Thesaurus UNESCO gọi tắt UT) được biên soạn từ năm 1973 - 1974, gắn liền với việc áp dụng máy tính điện tử vào công tác thông tin - tư liệu. Từ điển từ chuẩn của UNESCO gồm có 8.500 từ chuẩn phản ánh các lĩnh vực khoa học khác nhau như:

- Tổng loại
- Khoa học kỹ thuật
- Khoa học xã hội
- Con người và văn hoá

- Giáo dục
- Truyền thông
- Thông tin - thư viện - lưu trữ

Trong Từ điển từ chuẩn với mỗi thuật ngữ có thể bao gồm các chú giải về mặt:

- Thuật ngữ rộng hơn (Broader terms)
- Thuật ngữ hẹp hơn (Narrower terms)
- Thuật ngữ liên quan (Related terms)
- Dùng (Use)
- Dùng cho (Used for)
- Thuật ngữ cao nhất (Top terms)

Với một số từ chuẩn, có thể có thêm phần chú giải thuật ngữ (scope note) kèm theo.

Trong UT liệt kê bảng danh mục theo vần chữ cái các từ chuẩn có liên quan đến năm lĩnh vực chính mà UNESCO quan tâm như: giáo dục, khoa học kỹ thuật, khoa học xã hội, văn hoá và truyền thông.

UT cho đến nay vẫn là bộ từ điển từ chuẩn mang tính quốc tế duy nhất tồn tại và phổ biến rộng rãi.

Trong quá trình tra cứu, cán bộ thông tin phải căn cứ vào cơ sở dữ liệu cụ thể để xác định thư viện - cơ quan thông tin đó định từ khoá theo bộ từ khoá nào, trên cơ sở đó thể hiện yêu cầu tin cho sát. Tuy nhiên, dù cơ quan thông tin - thư viện có sử dụng bộ từ khoá nào đi chăng nữa, cán bộ tra cứu cũng phải linh hoạt để đáp ứng nhu cầu tin của người dùng tin.

Để tra cứu tin tự động hoá cần xây dựng phương trình tìm. Phương trình tìm cần hai yếu tố: Từ khoá và toán tử tìm.

Về phương diện cú pháp được thể hiện qua những quan hệ logic OR, AND, NOT (Liên quan đến vấn đề này sẽ được trình bày kỹ ở chương 4: Tìm tin tự động hoá).

2.2. Bộ máy tra cứu

Bộ máy tra cứu là tập hợp các công cụ và phương tiện cho phép tra tìm và cung cấp các tài liệu/ thông tin (dữ kiện, số liệu) phù hợp với diện đề tài bao quát của cơ quan thông tin - thư viện, đáp ứng yêu cầu tin của người dùng tin.

Trên cơ sở các hình thức lưu trữ thông tin, bộ máy tra cứu được phân chia thành:

- Bộ máy tra cứu thông tin truyền thống/ thủ công,
- Bộ máy tra cứu thông tin bán tự động hoá,
- Bộ máy tra cứu thông tin tự động hoá.

2.2.1. Bộ máy tra cứu thông tin truyền thống/ Thủ công

Bộ máy tra cứu truyền thống/ thủ công gồm: các bộ phiếu tra cứu thư mục, bộ phiếu tra cứu dữ kiện, hệ thống mục lục và kho tài liệu tra cứu.

2.2.1.1. Các bộ phiếu tra cứu thư mục

Là công cụ phản ánh các tài liệu/ thông tin về những vấn đề và đề tài mà cơ quan thông tin - thư viện phải phục vụ không phụ thuộc vào loại hình tài liệu và nơi bảo quản chúng.

+ Bộ phiếu tra cứu chính

Bộ phiếu tra cứu chính là một bộ phận chủ yếu, quan trọng nhất của bộ máy tra cứu. Bộ phiếu tra cứu chính có nhiệm vụ trả lời tất cả những yêu cầu tin theo diện bao quát để tài của cơ quan thông tin.

Đặc điểm của bộ phiếu tra cứu chính là nó bao quát toàn bộ những chuyên đề chính cũng như những chuyên đề có liên quan, kế cận; phản ánh các tài liệu không phụ thuộc vào các loại hình và nơi bảo quản chúng. Loại bộ phiếu này thường được xây dựng tại các thư viện chuyên ngành (thuộc các viện nghiên cứu : Viện nghiên cứu chè, Viện nghiên cứu cây lương thực - thực phẩm, Viện nghiên cứu Điện, Viện nghiên cứu Vật liệu xây dựng, ... hoặc tại các trung tâm thông tin như: Trung tâm thông tin Địa chất, Trung tâm thông tin thương mại,...

+ Các bộ phiếu tra cứu chuyên đề

Các bộ phiếu tra cứu chuyên đề được xây dựng để trả lời những yêu cầu tin theo những vấn đề, đề tài hẹp cụ thể nào đó được nhiều người quan tâm và có tính cấp bách (Tính thời sự).

Đặc điểm của các bộ phiếu chuyên đề là chúng phản ánh vấn đề một cách toàn diện nhất, đầy đủ nhất và mang tính thời sự rõ ràng.

Ví dụ:

Trong tình hình hiện nay, nạn ma túy có nguy cơ phát triển ở các trường học, để phát triển và ngăn chặn tệ nạn này các trường học đã quan tâm theo dõi và có biện pháp xử lý kịp thời. Thư viện nhiều trường đã xây dựng các hộp phiếu chuyên đề như "Tệ nạn ma túy và cách phòng chống" hoặc "Phòng và chống HIV/AIDS".

Đặc biệt các thư viện thuộc hệ thống thư viện công cộng thường xuyên tổ chức các hộp phiếu chuyên đề nhân dịp các ngày lễ lớn của

đất nước như: Ngày sinh nhật của Chủ tịch Hồ Chí Minh, Ngày Quốc khánh 2-9, các kì đại hội Đảng, ngày Giải phóng Thủ đô, Ngày chiến thắng lịch sử Điện Biên Phủ,..

Các thư viện đã thu thập thông tin từ các báo của Trung ương, địa phương và tạp chí định kỳ các nguồn tin này cung cấp thông tin một cách kịp thời, đầy đủ, và đảm bảo độ tin cậy.

+ *Các hộp phiếu thư mục*

Các hộp phiếu được tổ chức ở các thư viện rất đa dạng, trong đó phải kể đến bộ phiếu tra cứu theo loại hình tài liệu.

Các hộp phiếu tra cứu theo loại hình tài liệu giúp bạn đọc theo dõi thông tin về các loại hình tài liệu đã được nhập về thư viện một cách có hệ thống.

Nhiều thư viện và cơ quan thông tin đã xây dựng: *Hộp phiếu thư mục các bài trích báo, tạp chí.*

Ví dụ :

Trung tâm thông tin - Thư viện Trường Đại học Văn hoá Hà Nội đã xây dựng nhiều hộp phiếu thư mục bài trích (Báo, tạp chí, sách) như:

- Hộp phiếu thư mục bài trích về *Văn hoá nghệ thuật,*
- Hộp phiếu thư mục bài trích về *Thông tin - thư viện,*
- Hộp phiếu thư mục bài trích về *Bảo tồn bảo tàng,*
- Hộp phiếu thư mục bài trích về *Du lịch...*

Thư viện Trường Đại học Bách khoa Hà Nội đã xây dựng các hộp phiếu như:

- Hộp phiếu *Tạp chí ngoại văn,*
- Hộp phiếu *Tạp chí tiếng Việt,*

- Hộp phiếu *Luận án khoa học*,...

Ở hầu hết các thư viện và trung tâm thông tin thuộc các viện nghiên cứu như Viện khoa học kỹ thuật xây dựng, Viện vật liệu xây dựng, Viện khoa học kỹ thuật nông nghiệp,... đều xây dựng:

- Hộp phiếu các báo cáo khoa học và những công trình nghiên cứu: phản ánh những đề tài nghiên cứu đã được tiến hành, giúp các nhà nghiên cứu tham khảo và nghiên cứu không bị trùng lặp.

- Hộp phiếu catalo công nghiệp: giới thiệu các mẫu, mã sản phẩm của cơ quan, xí nghiệp Việt Nam và của nước ngoài.

- Hộp phiếu các công trình thiết kế thử nghiệm, ...

+ **Hộp phiếu tra cứu nhân vật/ Danh nhân**

Hộp phiếu này phản ánh về thân thế và sự nghiệp các nhân vật nổi tiếng trong các lĩnh vực khoa học - kỹ thuật, văn học - nghệ thuật, kinh tế, văn hoá,... trong nước và trên thế giới, giúp cho người dùng tin có điều kiện tìm hiểu một cách toàn diện, đầy đủ và nhanh chóng về các nhân vật đó.

Ví dụ:

Thư viện triều Nguyễn (thuộc Trung tâm bảo tồn Cố đô Huế - Tp. Huế) xây dựng: *Hộp phiếu các vua, chúa triều Nguyễn*,

Thư viện tỉnh Nghệ An xây dựng: *Hộp phiếu Danh nhân Nghệ Tĩnh*,

+ **Các hộp phiếu tra cứu về địa chí**

Hộp phiếu tra cứu về địa chí được xây dựng nhằm phản ánh về tài nguyên thiên nhiên, danh lam thắng cảnh, tiềm năng kinh tế, con người, ... của từng địa phương khác nhau.

Ví dụ: Thư viện Viện thông tin - bảo tàng địa chất khoáng sản đã tổ chức hợp phiếu địa chỉ về các khoáng sản có tại các địa phương, giúp cho cán bộ nghiên cứu tra cứu được nhanh chóng những thông tin cần thiết. Tuy nhiên, hầu hết các thư viện tỉnh đều đã tổ chức mục lục địa chỉ phản ánh thông tin một cách đầy đủ, toàn diện về địa phương mình.

+ Các bộ phiếu dữ kiện

Các bộ phiếu dữ kiện là những hệ thống tìm tin cho phép giải đáp các yêu cầu tin dữ kiện cụ thể, giúp người dùng tin không phải tra cứu tài liệu và quá trình rút ra từ những tài liệu có chứa dữ kiện cần thiết.

Đặc điểm của các bộ phiếu dữ kiện.

Đối tượng mô tả trong bộ phiếu dữ kiện không phải là một tài liệu trọn vẹn nào đó, mà là từng dữ kiện hoặc tập hợp những dữ kiện cần thiết được rút ra từ một/ nhiều tài liệu: những dữ kiện này liên quan với nhau trong một phạm vi một mẫu tìm/đối tượng hay quá trình.

Mỗi phiếu phản ánh một đối tượng

Thông tin trong phiếu dữ kiện phải được phân tích, so sánh và đánh giá nhằm bảo đảm độ tin cậy cho phép và tính chính xác cao.

Phiếu dữ kiện cần bảo đảm tính cập nhật và khi cần thiết có thể sửa đổi những dữ kiện.

Một số bộ phiếu dữ kiện thường được xây dựng trong cơ quan thông tin - thư viện.

- Bộ phiếu các sản phẩm, thiết bị mới.

Loại bộ phiếu này dùng để trả lời các yêu cầu về tính năng tác dụng, các đặc tính kỹ thuật, các đặc điểm thiết kế, giá cả, cơ quan xí nghiệp sản xuất.

Ví dụ:

Sản phẩm: Giấy da, đồ hộp, may mặc, đông lạnh, ...

Thiết bị: Thiết bị điện, thiết bị máy tính, thiết bị viễn thông,...

- *Bộ phiếu vật liệu*

Bộ phiếu vật liệu cung cấp dữ kiện về tính năng, tác dụng, lĩnh vực ứng dụng, cũng như: thành phần hoá học, cơ học, công nghệ chế tạo, giá thành, tính ưu việt so với các vật liệu tương tự, cơ quan thiết kế, cơ sở sản xuất.

Ví dụ:

Vật liệu xây dựng: sắt, thép, composit, ...

Vật liệu dệt: Bông, sợi, vải, ...

- *Bộ phiếu các quy trình công nghệ và các sáng kiến cải tiến*

Bộ phiếu này cung cấp những tin tức về những quy trình công nghệ tiên tiến, các trang thiết bị cơ giới hoá, tự động hoá đã được áp dụng ở những cơ sở nghiên cứu, sản xuất, về hiệu quả kinh tế của chúng và những sáng kiến, cải tiến kỹ thuật được đề xuất và áp dụng.

- *Bộ phiếu địa chỉ*

Được xây dựng để giúp đỡ các cơ sở sản xuất, các tổ chức, cơ quan biết được những địa chỉ cần thiết để giao dịch, mua, bán, xuất khẩu, nhập khẩu như:

- Địa chỉ khách hàng

- Địa chỉ cơ sở sản xuất các sản phẩm có liên quan, ...

Nhiều viện, trung tâm thông tin chuyên ngành đã xây dựng các cơ sở dữ liệu dữ kiện phục vụ cho cộng tác lưu trữ và tra cứu thông tin được đầy đủ, nhanh chóng, kịp thời.

2.2.1.2. Hệ thống mục lục thư viện

Hệ thống các mục lục thư viện là một bộ phận quan trọng của bộ máy tra cứu, là công cụ tra cứu truyền thống không thể thiếu được, nó phản ánh toàn bộ kho tài liệu gốc hiện có trong thư viện, giúp bạn đọc tra tìm tài liệu một cách thuận lợi và nhanh chóng. Hệ thống mục lục của thư viện gồm:

- **Mục lục chữ cái**

Mục lục chữ cái phản ánh kho sách theo tên tác giả hoặc tên sách (theo tiêu đề mô tả), vừa là phương tiện tuyên truyền sách, vừa là một công cụ tra cứu thư mục và sử dụng đơn giản.

Mục lục chữ cái trả lời các câu hỏi: Trong thư viện có bao nhiêu tác phẩm và cụ thể là những tác phẩm nào của một tác giả nào đó? Hoặc thư viện hiện đang lưu giữ những lần xuất bản nào của một tác phẩm.

- **Mục lục phân loại**

Mục lục phân loại cũng giống như mục lục chữ cái, là một công cụ tra cứu rất quan trọng trong thư viện.

Mục lục phân loại phản ánh kho tài liệu gốc của thư viện theo nội dung các ngành khoa học. Hiện nay trên thế giới nói chung và ở Việt Nam nói riêng tồn tại nhiều bảng phân loại khác nhau, khi tìm tin cần tham khảo xem thư viện đang sử dụng bảng phân loại nào để xử lý nội dung tài liệu (phân loại tài liệu) để khi tìm tin ta cũng phải sử dụng bảng phân loại đó để thể hiện yêu cầu tin/ mẫu tìm (ký hiệu phân loại).

Ví dụ:

Bảng phân loại BBK: Các ngành khoa học được sắp xếp theo vần chữ cái từ A - Z.

Bảng phân loại DDC: 000 - 999, Bảng phân loại UDC: 0 - 9 ...

- Mục lục chủ đề

Mục lục chủ đề phản ánh kho tài liệu gốc theo đề mục chủ đề. Cũng giống như khi phân loại tài liệu và xây dựng mục lục phân loại, thư viện sử dụng bảng đề mục chủ đề nào để định chủ đề và tổ chức mục lục chủ đề thì cũng theo bảng đó để tra cứu thông tin.

- Mục lục xuất bản phẩm định kỳ, tiếp tục

Giúp cán bộ thư viện, người dùng tin theo dõi và tra cứu các xuất bản phẩm định kỳ (các loại báo, tạp chí) thư viện thu thập được.

- Mục lục liên hợp

Mục lục liên hợp phản ánh tài liệu lưu giữ tại nhiều thư viện khác nhau trên địa bàn của một địa phương nào đó hoặc là một ngành nào đó.

- Mục lục địa chỉ

Mục lục địa chỉ giúp cho việc tra cứu một cách nhanh chóng và thuận lợi những tài liệu / thông tin đã được xuất bản có đề cập tới địa phương, vùng, miền nào đó.

2.2.1.3. Kho tài liệu tra cứu, ấn phẩm thông tin, thư mục

Số lượng và chất lượng của kho tài liệu này ảnh hưởng rất lớn tới hiệu quả phục vụ tra cứu của thư viện và cơ quan thông tin. Nếu tổ chức không tốt kho tài liệu tra cứu, cán bộ tra cứu có giỏi bao nhiêu cũng không thể trả lời được đầy đủ câu hỏi của người dùng tin.

- Kho tài liệu tra cứu

Kho tài liệu tra cứu là kho đặc biệt trong hệ thống kho sách của thư viện. Kho này được tổ chức trên cơ sở rút ra từ kho chính những

loại sách mang tính chất tham khảo, tra cứu thường xuyên được sử dụng đến, giúp cho cán bộ thông tin thư viện, người dùng tin tiện tra cứu hàng ngày tiết kiệm thời gian, công sức của họ.

Kho tài liệu tra cứu gồm:

+ **Những tài liệu có tính chất chỉ đạo**

Tài liệu mang tính chất chỉ đạo là các tài liệu của Đảng và Nhà nước (Văn kiện đại hội Đảng, các chỉ thị, nghị quyết, thông tư, quyết định của Chính phủ, ...) về phương hướng phát triển kinh tế, khoa học, văn hoá, xã hội,... của đất nước nói chung và của ngành nói riêng, phù hợp với diện phục vụ của thư viện, cơ quan thông tin.

Ví dụ:

✓ Văn kiện Đại hội Đảng lần thứ VI, VII, VIII, IX, Nghị quyết hội nghị TW, Điều lệ, ...

✓ Hiến pháp,

✓ Một số bộ luật, pháp lệnh : Bộ luật tố tụng hình sự, Luật xuất bản, Luật đầu tư nước ngoài, Pháp lệnh thư viện, Pháp lệnh lưu trữ,...

✓ Các tác phẩm kinh điển của chủ nghĩa Mác - Lênin (toàn tập, tuyển tập, tác phẩm riêng biệt, các tập chuyên đề của C.Mác, F. Englen, V. Lênin, Hồ Chí Minh,...).

+ **Bách khoa thư (Encyclopedias) và niên giám (Encyclopedias and Book of the years).**

Bách khoa thư cung cấp những kiến thức, tư liệu chính xác, cụ thể có hệ thống, về mọi vấn đề có liên quan tới đời sống chính trị, văn hoá, xã hội và khoa học. Hiện nay, ở Việt Nam chưa có bách khoa thư, song trên thế giới xuất hiện nhiều loại bách khoa thư từ thời cổ đại với các nhà bách khoa thư nổi tiếng như Aristotle (Ai Cập), Ploong the

Elder (La Mã), Alexander Neckam, Thomas Elyot, Ephraim Chambers (Anh),... Qua mỗi thời kì các Bách khoa thư lại hoàn chỉnh hơn về hình thức và phong phú hơn về nội dung, đáp ứng cho việc tra cứu thông tin của người dùng tin. Hiện nay, hầu hết các nước phát triển trên thế giới đều xuất bản bách khoa thư. Các ấn phẩm này đều được tái bản nhiều lần và ngày một hoàn thiện hơn. Ví dụ như: *Encyclopedia Britannica*: 16 lần (tới năm 2007); *Encyclopedia Americana*: 15 lần (tới năm 2006), Đại bách khoa thư (Liên Xô): 3 lần,...

Các bách khoa thư của thế giới được sử dụng rộng rãi ở các thư viện Việt Nam. Các loại bách khoa thư giúp cho cán bộ thư viện, bạn đọc tra cứu một cách thuận lợi và nhanh chóng về nhiều vấn đề mang nội dung tổng hợp.

Theo nội dung Bách khoa thư có thể chia ra làm nhiều loại:

- Bách khoa thư có nội dung tổng hợp: hầu hết các nước đều xuất bản loại bách khoa thư loại này.

Encyclopedia Britannica : (xuất bản lần 1: 1768 - 1771).

Lần xuất bản lần thứ 16 (2007) của *Encyclopedia Britannica* có 32 tập gồm: 44 triệu từ vựng; Hơn 65.000 bài báo; 15.000 bài báo về địa lý; 24.000 tranh ảnh minh họa, bản đồ; Hơn 23.000 tiểu sử nhân vật có từ Socrates đến Hồ Cẩm Đào được lấy từ hơn 4.000 nguồn tài liệu gốc. Ngoài hình thức in mang tính truyền thống trên giấy, *Encyclopedia Britannica* còn xuất bản dưới dạng điện tử (CD-ROM). *Encyclopedia Britannica* đổi mới, cập nhập thông tin kịp thời (mỗi năm đổi mới 10% lượng thông tin) bằng các phụ trương (niên giám):

Britanica book of the year. Encyclopedia Britannica, 19-,...200
-,... 2007 (Mỗi năm ra một phụ trương / niên giám).

Encyclopedia Americana

Nhiều thư viện ở các tỉnh phía Nam có bách khoa thư American. Encyclopedia American xuất bản lần thứ 15 (năm 2006) 30 tập gồm: 25 triệu từ vựng; 26.000 trang; 45.000 bài báo; 23.000 tranh ảnh minh họa; 1.200 bản đồ; 3.800 biểu đồ, ... Encyclopedia American cũng giống như Encyclopedia Britanica đổi mới, cập nhập thông tin kịp thời bằng các phụ trương (niên giám):

American Annual: An Encyclopedia of Current Events. N - Y, Encyclopedia Americana, 19-., 200- . Mỗi năm ra một phụ trương. Khi tra cứu cần sử dụng các Phụ trương/ niên giám vì chúng cung cấp thông tin mới và cập nhật.

- Bách khoa thư chuyên đề: cung cấp những kiến thức, tư liệu chính xác, cụ thể có hệ thống, chi tiết về vấn đề có liên quan tới một/ một số lĩnh vực khoa học. Loại bách khoa thư này được xuất bản nhiều ở các nước và Việt Nam.

Ví dụ:

- ✓ Bách khoa thư bệnh học,
- ✓ Bách khoa thư Quân sự,
- ✓ Bách khoa thư Hà Nội...

+ **Từ điển (Dictionaries)**

Cùng với bách khoa thư, từ điển cũng là công cụ tra cứu quan trọng cho đông đảo người dùng tin.

Từ điển gồm có một số loại chính như: từ điển bách khoa, từ điển ngôn ngữ, từ điển thuật ngữ, từ điển địa danh, từ điển tiểu sử/ nhân vật.

- *Từ điển bách khoa*

Từ điển bách khoa mang nội dung tổng hợp, cung cấp thông tin về tất cả các lĩnh vực khoa học, văn hóa, đời sống, xã hội. Tại một số nước, từ điển bách khoa được xuất bản khi chưa có điều kiện xuất bản bách khoa thư.

Ví dụ: Từ điển Bách khoa Việt Nam. 4T. (1990 - 2005) gồm 40.000 mục từ, có 1.200 nhà khoa học tham gia biên soạn.

- *Từ điển ngôn ngữ:*

Làm nhiệm vụ tập hợp đối chiếu và giải nghĩa các từ ngữ khái niệm của hai hay nhiều ngôn ngữ trên cơ sở một ngôn ngữ khác.

Ví dụ:

- ✓ Từ điển Anh - Việt,
- ✓ Từ điển Việt - Nga,
- ✓ Từ điển Hán - Việt...

- *Từ điển thuật ngữ*

Từ điển thuật ngữ tập hợp và giải thích ngắn gọn những từ, ngữ, khái niệm thuộc một lĩnh vực khoa học hoặc một ngành sản xuất.

Ví dụ:

- ✓ Từ điển Xã hội học,
- ✓ Từ điển Toán học,
- ✓ Từ điển Tiếng Việt, ...

- *Từ điển địa danh*

Từ điển địa danh cung cấp những thông tin chi tiết liên quan tới các địa danh văn hoá, lịch sử, du lịch, hoặc tới các địa phương, lãnh thổ của một quốc gia, một châu lục.

Ví dụ:

- ✓ Từ điển địa danh Văn hoá - Lịch sử Việt Nam,
- ✓ Từ điển Hà Nội địa danh ...

+ *Sách tra cứu, sổ tay kỹ thuật (handbooks)*

Sách tra cứu cung cấp những kiến thức cơ bản nhất, cần thiết nhất về một ngành hoặc một lĩnh vực sản xuất cụ thể. Theo nội dung, có thể chia sách tra cứu thành ba nhóm: sách tra cứu có nội dung tổng hợp, sách tra cứu chuyên đề và sách tra cứu về tiêu chuẩn, quy tắc, điều lệ.

• *Sách tra cứu kỹ thuật*

Sách tra cứu kỹ thuật mang tính chất tổng hợp thường được xuất bản dưới dạng cẩm nang, sổ tay tra cứu để phục vụ cho một đối tượng cụ thể, nhằm cung cấp những kiến thức cô đọng và chủ yếu về từng ngành khoa học hoặc lĩnh vực sản xuất.

Ví dụ:

- ✓ Sách tra cứu tóm tắt về vật lý,
- ✓ Sổ tay kỹ sư thiết kế chế tạo máy,
- ✓ Sổ tay thiết bị khoa học,
- ✓ Sổ tay phóng viên.

▪ *Sách tra cứu chuyên đề*

Sách tra cứu chuyên đề tập hợp tư liệu về từng vấn đề, đề tài cụ thể của từng lĩnh vực hoạt động sản xuất: về nguyên vật liệu, máy móc thiết bị, về tổ chức quản lý sản xuất, về phương pháp canh tác,...

Sách tra cứu chuyên đề có thể trình bày những vấn đề rộng/
Những đề tài hẹp: liên quan đến nhiều ngành sản xuất.

Ví dụ:

- ✓ Sách tra cứu về ổ bi, khớp nối,
- ✓ Bảng tra số liệu kết cấu thép,
- ✓ Chỉ dẫn về phân bón,
- ✓ Bảng tra chữ nôm.
- *Sách tra về tiêu chuẩn, quy tắc, thể lệ*

Sách tra về tiêu chuẩn, quy tắc, thể lệ là loại tài liệu tập hợp yêu cầu, điều kiện cơ bản về một lĩnh vực hoạt động nào đó đã được cơ quan có thẩm quyền ban hành và sử dụng.

Ví dụ:

- ✓ Văn bản pháp quy về thư viện.
- ✓ Thể lệ giao thông đường biển.
- ✓ Điều lệ trật tự an toàn giao thông vận tải đường bộ.
- + *Niên biểu, lịch biểu, niên giám*

Loại tài liệu này cung cấp các số liệu, kiến thức cho NDT về các lĩnh vực kinh tế, văn hoá, xã hội trong những khoảng thời gian nhất định.

Ví dụ:

- ✓ Niên giám Thống kê (Xuất bản hàng năm).
- ✓ Niên giám công thương Hà Nội - Hải Phòng - TP. Hồ Chí Minh 1997-1998.
- ✓ Nông lịch ... (Xuất bản hàng năm).

Đáng kể là trong loại tài liệu này là **Almanach**: theo nghĩa tiếng Việt Almanach có nghĩa là là niên lịch, niên giám, được xuất bản trong năm. Kiến thức trong Almanach tổng hợp như một cuốn bách khoa thư thu nhỏ.

Ví dụ:

- ✓ Almanach - Sách lịch phổ thông (Xuất bản hàng năm).
- ✓ Almanach - Những nền văn minh thế giới ...

+ **Nguồn tra cứu về các cơ quan, tổ chức, tiểu sử và địa lý**

- *Nguồn tra cứu về các cơ quan, tổ chức: Sách dẫn (Directory, guide)*

Sách dẫn về các cơ quan, tổ chức là một trong những nguồn tra cứu chính. Nguồn tra cứu về các tổ chức - cơ quan là danh mục tên của các cơ quan - tổ chức sắp xếp theo trật tự chữ cái thuộc lĩnh vực hoạt động, sau đó theo trật tự chữ cái tên của các cơ quan: bao gồm thông tin chi tiết, đầy đủ tên cơ quan/ chi nhánh/ đại diện, địa chỉ, số điện thoại, tên của thủ trưởng cơ quan, chức năng, nhiệm vụ của cơ quan, về hành chính tổ chức như: lịch sử hình thành cơ quan/ tổ chức, lĩnh vực hoạt động, các loại dịch vụ và các tài liệu có liên quan đã được in ấn.

Ví dụ:

- ✓ Các thư viện Việt Nam.
- ✓ *Directory of special libraries in Australia.*
- ✓ *World of learning.*
- ✓ *Press, radio and TV guide.*
- *Nguồn tra cứu về tiểu sử (Dictionary of biography)*

Nguồn tra cứu về tiểu sử cung cấp những tin tức ngắn gọn về cuộc đời và sự nghiệp của các nhà hoạt động chính trị, khoa học, văn hoá, xã hội nổi tiếng trong nước và trên thế giới. Nhiều nguồn tra cứu thông tin rất chi tiết đầy đủ về cuộc đời và sự nghiệp của các nhân vật đó như: cuộc đời riêng (liệt kê các mốc thời gian về tiểu sử một cách chi tiết nhất, cách phát âm tên theo ngôn ngữ gốc, sở thích, nghề

nghiệp, thành viên của các hiệp hội, công trình đã công bố, có đăng kèm theo chân dung hoặc ảnh của nhân vật. Thông thường loại tài liệu này được xuất bản dưới dạng từ điển nhân vật.

Ví dụ:

- ✓ Từ điển nhân vật lịch sử Việt Nam.
- ✓ Tác gia văn học Thăng Long.
- ✓ 330 danh nhân thế giới, ...

Nhiều nước, nhiều tổ chức quốc tế tiến hành xuất bản các ấn phẩm định kỳ, ấn phẩm tiếp tục hoặc các thư mục tiểu sử (Biography index, biography directories) về các nhân vật và chính khách trong nước và trên thế giới:

Ví dụ:

- ✓ International who's who (Xuất bản tại Anh).
- ✓ Who's who in America (Xuất bản tại Mỹ).
- ✓ Australian dictionary of biography (Xuất bản tại Ôxtralia), ...
- *Nguồn tra cứu về địa lý*

Đây là nguồn tra cứu thông tin về các địa danh bao gồm: các thành phố, thị trấn, làng, xã, sông ngòi, núi và hồ. Nguồn tra cứu về địa lý thường được xuất bản dưới dạng:

○ **Bản đồ:**

Là loại tài liệu mô tả sơ lược về kích thức bề mặt của thế giới, châu lục hoặc một địa phương cụ thể. Có nhiều loại bản đồ được xuất bản như: bản đồ địa lí, bản đồ chính trị, bản đồ kinh tế, bản đồ chuyên đề, ...

Ví dụ:

- ✓ Bản đồ khoáng sản Việt Nam.
- ✓ Bản đồ thế giới.
- ✓ Bản đồ Hà Nội.
- ✓ Tập bản đồ (Atlases):

Các loại bản đồ và những chi dẫn giúp cho người dùng tin tra tìm những thông tin cần thiết. Cũng như bản đồ, tập bản đồ có nhiều loại: tập bản đồ thế giới, tập bản đồ quốc gia, tập bản đồ lịch sử, tập bản đồ chi dẫn đường và phố, tập bản đồ kinh tế,... Ví dụ: Atlas - Bản đồ thế giới (Dùng cho học sinh phổ thông)

○ **Sách tra cứu địa lí :**

Loại tài liệu tra cứu này liệt kê các địa danh và những địa phương trực thuộc. Sách tra cứu địa lí là nguồn tra cứu quan trọng vì chúng cung cấp những thông tin liên quan đến sự kiện, dữ liệu: vị trí địa lí, diện tích, dân số, cách phát âm đúng... liên quan tới địa danh, mô tả chúng ở thời hiện tại cũng như cung cấp những thông tin mang tính lịch sử.

Ví dụ:

- ✓ Từ điển Địa danh Văn hoá Việt Nam.
- ✓ Từ điển Địa danh Lịch sử - Văn hoá Việt Nam, ...

Các nguồn tra cứu địa lí khác: Bách khoa thư, Niên giám, Từ điển bách khoa... cũng có thể tra cứu thông tin liên quan tới các địa danh.

- **Ấn phẩm thông tin - thư mục**

Ấn phẩm thông tin - thư mục cung cấp cho người dùng tin và cán bộ thông tin - thư viện những tin tức về nguồn tài liệu xuất bản

trong nước và trên thế giới, không phân biệt tài liệu đó có trong thư viện hay không.

Ấn phẩm thông tin thư mục gồm: các bản thư mục, tạp chí tóm tắt (abstracts), thông tin nhanh (Express informations), thông tin tín hiệu (Signal informations), tổng quan... là tài liệu cấp 2, cấp 3 do các cơ quan thông tin thư viện biên soạn trên cơ sở phân tích, xử lý, tổng hợp tài liệu cấp 1 (tài liệu gốc).

Trên thế giới có nhiều ấn phẩm thông tin thư mục nổi tiếng, có uy tín được xuất bản và phổ biến rộng rãi trên thế giới cũng như ở Việt Nam:

Chemical Abstracts (Mỹ).

Biological Abstracts (Mỹ).

Agrindex (FAO)

Bulletin Signalitique (Pháp).

Реферативный журнал (Nga).

Japanese Science Reviews (Nhật).

Tại Việt Nam các cơ quan thông tin - thư viện cũng biên soạn và xuất bản nhiều ấn phẩm thông tin thư mục như:

❖ *Trung tâm Thông tin Khoa học và Công nghệ Quốc gia:*

- Thông báo sách mới.
- Tạp chí tóm tắt tài liệu khoa học và công nghệ Việt Nam.
- Tổng quan Khoa học - Công nghệ - Kinh tế,...

❖ *Thư viện Quốc gia xuất bản:*

- Thư mục Quốc gia Việt Nam.
- Thông tin Văn hoá nghệ thuật,...

❖ *Viện thông tin khoa học Xã hội:*

- Thư mục sách ngoại văn.

- Thông báo sách mới...

❖ *Trung tâm thông tin Cục sở hữu trí tuệ:*

- Công báo sở hữu trí tuệ. 2 tập.

Ngoài ra, các thư viện tỉnh, thành phố còn biên soạn các thư mục địa chỉ.

Với sự phong phú về nội dung và đa dạng về hình thức, tài liệu tra cứu đóng vai trò quan trọng trong việc tổ chức bộ máy tra cứu và phục vụ tra cứu cho người dùng tin. Nó không những chỉ giúp tra cứu tài liệu/ thông tin cần thiết mà còn là cơ sở để tìm hiểu về những khái niệm, thuật ngữ khoa học và nhiều vấn đề cơ bản khác nhằm đáp ứng đầy đủ hơn những yêu cầu về thông tin và tra cứu của bạn đọc.

2.2.2. Bộ máy tra cứu bán tự động hoá

Bộ máy tra cứu bán tự động được xây dựng trên nguyên tắc lưu trữ thông tin theo hình thức của tài liệu hoặc theo nội dung trên các phiếu lỗ mép và phiếu lỗ soi.

2.2.2.1. Phiếu lỗ mép

Phiếu lỗ mép là một hệ thống lưu trữ thông tin theo nguyên tắc "xếp theo tài liệu", nghĩa là mỗi tài liệu được trình bày trên một phiếu. Phiếu này là thành phần chính của bộ nhớ. Phiếu gồm có hai vùng:

- **Vùng giữa phiếu:** dùng để mô tả thư mục.

- **Vùng xung quanh mép phiếu:** là vùng đục lỗ theo một mã số đã quy định. Người ta có thể ghi dưới dạng các lỗ khuyết những đặc

trung nội dung và khái niệm của tài liệu, nhằm giúp cho việc tra cứu một cách nhanh chóng. Khi cần tra cứu tài liệu về một vấn đề nào đó, người tìm tin sử dụng que xiên vào lỗ tương ứng từng phiếu một với đặc trưng định tìm, rồi cho tập phiếu rơi tự do. Những phiếu có bắm lỗ khuyết sẽ rơi xuống - đó chính là kết quả của sự lựa chọn.

Nếu một đặc trưng được trình bày bằng nhiều lỗ và nếu muốn tra cứu theo nhiều đặc trưng, thì phải lập lại phương pháp trên cho mỗi lỗ.

Phiếu lỗ mép có thể được làm bằng bìa cứng, bìa mềm, giấy Bristol, có kích thước từ 75 x 125mm đến 210 x 297mm. Các lỗ có đường từ 2,5 - 3mm và khoảng cách giữa chúng là 5,6 mm hoặc 6,35mm.

Mã số đăng ký:

Mã số có thể là Số hoặc Chữ cái, tùy theo mã số được tạo nên trong quá trình xây dựng phiếu. Có nhiều loại mã số được sử dụng trong quá trình mã hoá tài liệu, có những cách mã hoá cho phép đăng ký một số lượng tin lớn và có khả năng sử dụng một cách rất mềm dẻo và linh hoạt. Một vài mã số thông dụng như 1.2.4.7., mã tam giác, mã chữ cái,...

2.2.2.2. Phiếu lỗ soi

Phiếu lỗ soi được xây dựng ngược lại với phiếu lỗ mép, trên cơ sở mỗi phiếu ứng với một từ khoá hoặc từ chuẩn đặc trưng cho nội dung tài liệu. Mỗi lỗ tương trưng cho một số đăng ký của tài liệu trong hệ thống. Để ghi một tài liệu vào bộ nhớ ta phải đục lỗ ở vị trí dành cho nó trong những phiếu ứng với khái niệm mà tài liệu đó có đề cập tới.

Khi cần tra cứu tài liệu có liên quan tới một từ khoá, từ chuẩn nào đó, ta chỉ việc đặt chồng các phiếu có cùng đặc trưng trước nguồn sáng. Những tài liệu phù hợp với yêu cầu sẽ là những phiếu có những vị trí đục lỗ trên toàn bộ những phiếu chồng lên nhau và vì vậy mà ánh sáng có thể xuyên qua được.

Phiếu lỗ soi được làm bằng giấy bristol nhẹ có nhiều loại kích thước khác nhau. Trên mỗi phiếu có một mạng các vị trí được xây dựng bằng phương pháp tọa độ. Mỗi vị trí ứng với một số đăng ký của tài liệu đã được gán cho từ trước. Số lượng các vị trí trên mỗi phiếu ứng với số lượng của tài liệu có thể đưa vào bộ nhớ. Dung lượng của mỗi phiếu có thể từ 5000 tài liệu (như mẫu phiếu SPHINXO) hoặc 14.000 tài liệu (như mẫu phiếu SELECTO).

Trong mẫu SELECTO, mỗi phiếu có kích thước 150 x 200mm. Phiếu bao gồm các ô vuông đánh số theo hoành độ từ 0 đến 99 và theo tung độ từ 0 đến 139. Như vậy sẽ có 14000 vị trí ứng với mỗi ô vuông.

Lưu trữ thông tin:

Mỗi tài liệu trước khi đưa vào hệ thống đều được ghi một số đăng ký theo thời gian.

Ví dụ: Muốn đục lỗ tài liệu "*Bệnh lao phổi ở miền núi Việt Nam*" xuất bản năm 1987, ta lấy các phiếu tương ứng với các từ khoá:

- Lao phổi
- Miền núi
- Việt Nam
- Năm 1987

Nếu tài liệu đưa vào hệ thống có số thứ tự 5498 thì mỗi phiếu trích ra sẽ được đục lỗ ở ô vuông có tọa độ:

$$X = 54, Y = 98.$$

Việc đục lỗ có thể thực hiện bằng máy đục lỗ điện tử.

Việc đọc được thực hiện bằng cách sử dụng một khung chiếu sáng, cho phép nhìn thấy dễ dàng các ô có lỗ chung với nhau trên các phiếu chồng nhau. Mỗi lần có thể chồng khoảng 30 phiếu lên nhau, tuy nhiên điều này ít gặp trong thực tế.

Hiện nay cùng với sự phát triển rất mạnh mẽ của công nghệ thông tin, trong hoạt động của các cơ quan thông tin - thư viện hầu như không còn áp dụng lưu trữ và tra cứu thông tin bán tự động nữa, mà hoặc họ sử dụng hình thức lưu trữ và tra cứu thông tin truyền thống, hoặc sử dụng hình thức lưu trữ và tra cứu thông tin tự động hoá. Tổ chức hình thức lưu trữ và tra cứu thông tin bán tự động tồn tại không lâu, nay nó chỉ còn mang tính lịch sử.

2.2.3. Bộ máy tra cứu thông tin tự động hoá

Ngày nay với sự bùng nổ thông tin, người ta khó có thể chỉ dựa những công cụ tra cứu thủ công (như các bộ phiếu) để tìm kiếm nhanh chóng và đầy đủ những thông tin cần thiết. Để giải quyết tình trạng này người ta đã xây dựng bộ máy tra cứu thông tin tự động hoá. Bộ máy tra cứu thông tin tự động hoá thường được gọi là **hệ thống tìm tin tự động hoá**.

Thuật ngữ "**hệ thống tìm tin**" nói chung được sử dụng để chỉ một hệ thống mà nó lưu trữ thông tin và cho phép tìm lại thông tin đã lưu trữ. Trong một chừng mực nào đó và theo quan niệm truyền thống,

một tủ mục lục thư viện, một tạp chí tóm tắt (như Chemical Abstracts hoặc Реферативный журнал), tệp thư mục... có thể được coi như một hệ thống tìm tin thủ công. Tuy nhiên, theo quan niệm hiện đại, hệ thống tìm tin phải là hệ thống được tin học hoá có sử dụng máy tính điện tử, có phần mềm lưu trữ và tìm tin trên các cơ sở dữ liệu.

Hệ thống tìm tin tự động hoá là *hệ thống có khả năng lưu trữ, tìm lại và bảo trì thông tin được tin học hoá. Hệ thống tìm tin tự động hoá bao gồm phần cứng, phần mềm, các công cụ ngữ nghĩa phục vụ xử lý thông tin và yêu cầu tin, các cơ sở dữ liệu để người tìm tin truy cập và tìm lại được những thông tin đã lưu trữ trong đó và nhân viên làm việc với hệ thống.*

2.2.3.1. Thành phần hệ thống tìm tin tự động hoá

Những thành phần của hệ thống tìm tin tự động hoá có quan hệ tương tác với nhau và mỗi thành phần được thiết kế để thực hiện những chức năng nhất định. Hệ thống tìm tin tự động hoá có thể là hệ thống đơn giản chỉ gồm máy tính điện tử độc lập, phần mềm và cơ sở dữ liệu. Nhưng hệ thống tìm tin tự động hoá có thể là hệ thống rất phức tạp với nhiều thành phần hợp thành như trường hợp các dịch vụ thông tin trực tuyến.

Về tổng quát, hệ thống tìm tin tự động hoá bao gồm những thành phần cơ bản như sau:

Phương tiện kỹ thuật,

Hệ thống công cụ xử lý logic, ngữ nghĩa,

Cơ sở dữ liệu,

Nhân viên làm việc với hệ thống.

- **Phương tiện kỹ thuật:** là thành phần rất quan trọng của hệ thống tìm tin và bao gồm phần cứng, phần mềm, các phương tiện viễn thông.

+ **Phần cứng** bao gồm các máy tính điện tử, các thiết bị ngoại vi, các bộ nhớ (lưu trữ thông tin) và các thiết bị mạng. Máy tính điện tử là một thiết bị điện tử xử lý rất nhanh thông tin đưa vào dưới sự điều khiển của các chương trình lưu trữ trong bộ nhớ. Máy tính điện tử có những khả năng rất mạnh trong quản lý, xử lý và trao đổi dữ liệu.

+ **Bộ nhớ ngoài** là thiết bị để lưu trữ chương trình và thông tin. Trong hệ thống tìm tin tự động hoá người ta sử dụng các bộ nhớ ngoài như băng từ, đĩa từ và đĩa quang.

Băng từ là băng nhựa có phủ lớp vật liệu có khả năng nhiễm từ. Mặt băng rộng 12,7mm được chia thành 7 hay 9 rãnh từ, trong đó rãnh cuối cùng được sử dụng để kiểm tra. Một băng từ dài khoảng 730m có khả năng lưu trữ đến 46 triệu ký tự. Băng từ chủ yếu được sử dụng để sao lưu dữ liệu.

Đĩa từ là một đĩa bằng kim loại hoặc chất dẻo, trên mặt đĩa có phủ lớp vật liệu có khả năng nhiễm từ. Việc lưu trữ thông tin trên đĩa từ được thực hiện trên cơ sở công nghệ điện tử. Đĩa từ có hai loại là đĩa cứng và đĩa mềm.

Đĩa quang (Optical disc) là vật mang tin quang học. Đĩa quang được tạo ra từ đầu những năm 80 của thế kỷ XX. Đĩa quang được sử dụng nhiều trong hoạt động thông tin thư viện là đĩa CD-ROM (tên viết tắt tạo từ từ tiếng Anh Compact Disc - Read Only Memory). Đĩa CD-ROM có dung lượng lưu trữ khoảng 650 Mb và được sử dụng để phổ biến các cơ sở dữ liệu hoặc các tài liệu toàn văn. Để đọc được đĩa CD-ROM, máy tính cần có thiết bị đặc biệt là đầu đọc CD-ROM.

Hiện nay có những thiết bị có khả năng đặt vào đó đồng thời nhiều đĩa CD-ROM và đọc nhiều đĩa CD-ROM. Đó là các tháp CD-ROM (CD-ROM Tower) và CD-ROM Jukebox.

+ *Phần mềm*: bao gồm những phần mềm hệ thống (như hệ điều hành) và phần mềm ứng dụng (như chương trình quản trị mạng, chương trình quản trị cơ sở dữ liệu, chương trình viễn thông). Mạng máy tính đã trở thành một công cụ chiến lược trong hoạt động của hầu hết các tổ chức, doanh nghiệp. Sự phát triển của mạng máy tính gắn liền với sự phát triển của công nghệ máy tính điện tử, công nghệ thông tin và công nghệ viễn thông. *Mạng máy tính là một tập hợp các máy tính được kết nối với nhau bởi các đường truyền vật lý theo một kiến trúc nào đó.* Những thiết bị mạng, ngoài các máy tính điện tử và các thiết bị viễn thông có thể bao gồm các (card) mạng, các đường truyền dữ liệu (cáp đồng, cáp quang), bộ tập trung (HUB), bộ định tuyến (Router), bộ chuyển mạch (Switch).

Ngày nay với sự phát triển của máy tính điện tử, công nghệ mạng máy tính và viễn thông đã tạo ra các mạng thông tin máy tính. Các thiết bị viễn thông rất đa dạng. Một trong những thiết bị viễn thông được sử dụng rộng rãi trong khai thác các mạng thông tin máy tính là Modem. Đây là thiết bị cho phép kết nối máy tính điện tử với mạng thông tin máy tính thông qua đường điện thoại.

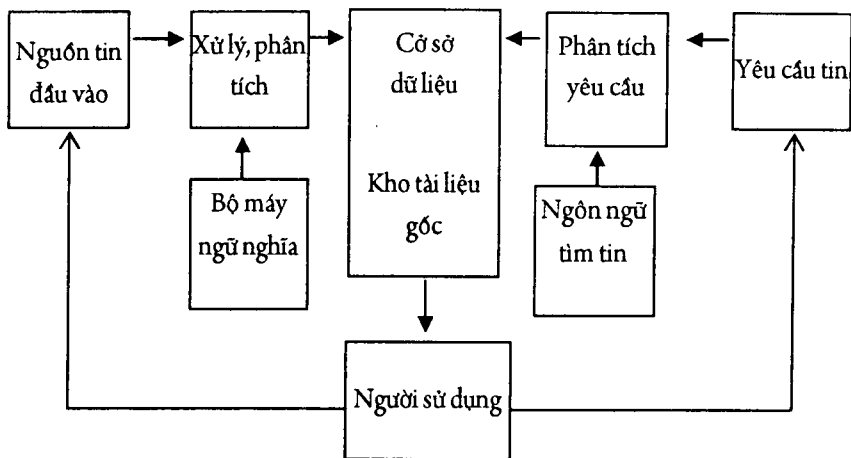
- **Hệ thống công cụ xử lý logic, ngữ nghĩa** bao gồm: tập hợp các công cụ dùng cho công tác xử lý hình thức và xử lý nội dung thông tin như các quy định về mô tả tài liệu (quy tắc mô tả thư mục AACR2, ISBD,...), các ngôn ngữ tìm tin (như các ký hiệu phân loại, từ điển từ chuẩn, đề mục chủ đề, bộ từ khoá có kiểm soát,...), các khổ mẫu dữ

liệu (như MARC 21, UNIMARC, CCF,...), các quy tắc xử lý thông tin, nhập dữ liệu... Những công cụ xử lý logic ngữ nghĩa là thành phần đóng vai trò cực kỳ quan trọng trong việc phân tích, trình bày và tổ chức lưu trữ thông tin phục vụ cho việc tìm kiếm lại thông tin đã được lưu trữ. Những công cụ của bộ máy logic ngữ nghĩa còn có vai trò quan trọng trong việc phân tích yêu cầu tin và xây dựng biểu thức tìm để đưa vào hệ thống tìm tin.

- **Cơ sở dữ liệu (CSDL)** là tập hợp có cấu trúc các các dữ liệu về đối tượng được quản lý thông tin. CSDL là một thành phần cơ bản của hệ thống tìm tin tự động hoá. Nhiều dịch vụ thông tin có thể lưu giữ trong hệ thống máy tính của mình nhiều CSDL tạo thành ngân hàng dữ liệu (NHDL), còn gọi là những nhà bán hàng dịch vụ trực tuyến (vendor) hoặc các dịch vụ tìm tin (search services). Hệ thống tìm tin có thể có nhiều loại CSDL khác nhau, từ cơ sở dữ liệu thư mục (CSDL TM), CSDL chỉ dẫn đến các CSDL dạng dữ kiện (CSDL DK: số, văn bản-số), tính chất, và CSDL toàn văn.

Cơ sở dữ liệu được tạo ra, quản lý và khai thác bằng một hệ thống phần mềm gọi là quản trị cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu cho phép người sử dụng có thể thực hiện những chức năng cơ bản sau đây trong việc quản lý dữ liệu:

- Mô tả dữ liệu.
- Cập nhật và bảo trì dữ liệu.
- Tìm kiếm dữ liệu lưu trữ.



Hình 1. Mô hình hoạt động đơn giản hệ thống tìm tin

Đối tượng quản lý thông tin trong cơ sở dữ liệu rất đa dạng. Đó có thể là các thông tin về những bài tạp chí, sách, tuyển tập, luận án tiến sỹ, sáng chế, tiêu chuẩn, hồ sơ cá nhân, hồ sơ công ty, số liệu thống kê, công thức hoá học, thành phần và cấu tạo các chất, v.v...

- **Con người:** làm việc với hệ thống gồm những nhóm cán bộ chuyên môn khác nhau có trách nhiệm thiết kế, xây dựng, vận hành, bảo trì, khai thác hệ thống và tìm tin để đáp nhu cầu tin của khách hàng hoặc của bản thân.

Mô hình hoạt động đơn giản của một hệ thống tìm tin có thể được trình bày trong hình 1.

2.2.3.2. Phân loại hệ thống tìm tin tự động

Người ta có thể phân loại hệ thống tìm tin tự động hoá theo mục đích cung cấp dịch vụ (cho nhu cầu nội bộ hoặc đáp ứng nhu cầu bên ngoài), theo người cung cấp, theo loại hình thông tin của hệ thống.

- Phân loại theo mục đích cung cấp dịch vụ.

Theo mục đích cung cấp dịch vụ, có thể phân biệt hai loại hệ thống tìm tin tự động hoá: hệ thống tìm tin nội bộ (in-house system) và hệ thống tìm tin kinh doanh.

+ *Hệ thống tìm tin nội bộ*: được xây dựng nhằm đáp ứng nhu cầu tìm tin của người dùng tin thuộc tổ chức nhất định. Ví dụ mục lục thư viện truy cập trực tuyến (OPAC) là một dạng hệ thống tìm tin tự động hoá nội bộ vì nó được xây dựng để đáp ứng nhu cầu tra cứu tìm tin của bạn đọc của thư viện.

+ *Hệ thống tìm tin tự động hoá kinh doanh*: hoạt động dựa trên cơ chế thị trường.

Những hệ thống này cung cấp dịch vụ truy cập đến các cơ sở dữ liệu hoặc ngân hàng dữ liệu cho người sử dụng. Ví dụ, cho hệ thống này là các dịch vụ thông tin trực tuyến. Người sử dụng thanh toán phí dịch vụ căn cứ vào mức độ khai thác sử dụng dịch vụ (như theo câu hỏi, theo số biểu ghi được xem, được tải về, thời gian truy cập, loại hình cơ sở dữ liệu truy cập). Những hệ thống này thường được gọi là các dịch vụ tìm tin hoặc dịch vụ trực tuyến.

- Phân loại theo phương thức truy cập

Theo phương thức truy cập cơ sở dữ liệu người có thể chia hệ thống tìm tin thành:

- + Hệ thống tìm tin cục bộ.
- + Hệ thống tìm tin trực tuyến.
- + Hệ thống tìm tin trên CD-ROM.
- + Hệ thống thông tin toàn cầu (Internet).

Bảng 1. Đặc điểm của một số hệ thống tìm tin

Hệ thống tìm tin	Đặc tính người sử dụng	Môi trường sử dụng	Chức năng	Công nghệ truy cập
Hệ thống tìm tin trực tuyến	Chuyên gia tìm tin; Người quản trị; Người dùng tin đầu cuối.	Các trường đại học, các trung tâm thông tin, thư viện.	Tìm tin, cung cấp thông tin thư mục, dữ kiện, toàn văn.	Mạng máy tính; Kênh thuê bao riêng; Internet.
CD-ROM	Người dùng tin đầu cuối Chuyên gia tìm tin.	Các trường đại học, các trung tâm thông tin, thư viện.	Các cơ quan nghiên cứu KH Tìm tin, cung cấp thông tin thư mục, dữ kiện, toàn văn.	Đồ họa, trên nền Windows, đa phương tiện.
Internet	Người dùng tin đầu cuối, Công chúng.	Trường đại học, trung tâm thông tin - thư viện; cơ quan; nghiên cứu KH; Gia đình.	Thông tin đại chúng, toàn văn Phương tiện liên lạc (email).	Mạng viễn thông, TCP/IP, Tựa Web (Web based).

Hệ thống tìm tin cục bộ cung cấp khả năng truy cập đến các cơ sở dữ liệu lưu trữ trên máy tính điện tử tại chỗ trong khi hệ thống tìm tin trực tuyến cung cấp khả năng truy cập đến các cơ sở dữ liệu lưu trữ trên máy tính điện tử ở xa thông qua mạng máy tính và hệ thống viễn thông. Cùng với sự phát minh ra CD-ROM như một loại hình vật mang tin quang học, rất nhiều hệ thống tìm tin trực tuyến và các nhà sản xuất cơ sở dữ liệu đã đưa CSDL lên CD-ROM và phổ biến chúng cho người sử dụng theo cơ chế kinh doanh. Những nguyên lý và kỹ thuật tìm tin trên CD-ROM, trên hệ thống cục bộ và trực tuyến về cơ bản không khác biệt. Điểm khác biệt giữa hệ thống tìm tin trực tuyến với những hệ thống khác là sự có mặt của mạng viễn thông để hỗ trợ truy cập hệ thống theo phương thức từ xa.

Gần đây với sự phát triển nhanh chóng của công nghệ thông tin và công nghệ viễn thông đã tạo ra một mạng thông tin toàn cầu - mạng Internet. Đặc biệt, với sự phát minh ra World Wide Web (còn gọi tắt là Web hoặc WWW), khu vực hoạt động của hệ thống tìm tin trực tuyến đã mở rộng. Internet và Web đã giúp làm cho thông tin trên mạng có thể được truy cập bởi bất cứ ai nếu máy tính của họ được kết nối với Internet. Điều này đã dẫn đến sự hình thành khái niệm hệ thống thông tin toàn cầu. Khai thác hệ thống này được trình bày ở phần tìm tin trên Internet. Một số đặc điểm của những hệ thống tìm tin tự động hoá được trình bày trong bảng 1.

Bảng 2. Phân loại hệ thống tìm tin theo thế hệ

Thế hệ	Loại thông tin	Công nghệ sử dụng
Hệ thống tìm tin thế hệ I.	Thông tin thư mục.	<ul style="list-style-type: none"> - Dòng lệnh; - Sử dụng hệ điều hành DOS hoặc mạng; - Người sử dụng là cán bộ chuyên nghiệp, đòi hỏi có nhiều kỹ năng. - Ví dụ: các dịch vụ trực tuyến.
Hệ thống tìm tin thế hệ II.	Thông tin thư mục và toàn văn.	<ul style="list-style-type: none"> - Sử dụng hệ thống thực đơn; - Sử dụng đơn giản hơn. <p>Người sử dụng là chuyên nghiệp hoặc người dùng đầu cuối.</p> <ul style="list-style-type: none"> - Hệ thống trên CD-ROM.
Hệ thống tìm tin thế hệ III.	Thông tin đa phương tiện (Multimedia).	<ul style="list-style-type: none"> - Sử dụng giao diện đồ họa; - Sử dụng chuột (Click); - Sử dụng đơn giản; - Dùng trình duyệt Web để khai thác; - Các hệ thống trên CD-ROM và trên Internet.

- Phân loại theo thế hệ của hệ thống tìm tin

Theo thế hệ của hệ thống, người ta phân hệ thống tìm tin tự động hoá thành ba loại (xem bảng 2).

- Hệ thống tìm tin thế hệ thứ nhất: là hệ thống tìm tin thư mục.
- Hệ thống tìm tin thế hệ thứ hai: là hệ thống tìm tin toàn văn.
- Hệ thống tìm tin thế hệ thứ ba: là hệ thống tìm tin đa phương tiện.

Tuy nhiên, việc phân loại cũng có tính chất tương đối vì một hệ thống có thể thực hiện nhiều chức năng.

2.2.3.3. Cơ sở dữ liệu

Cơ sở dữ liệu là một tập hợp có cấu trúc các dữ liệu về đối tượng được quản lý, lưu trữ đồng thời trên các vật mang tin của máy tính điện tử và được quản lý theo phương thức thống nhất nhằm giúp cho việc tìm tin được dễ dàng và nhanh chóng.

Cơ sở dữ liệu là tập hợp theo phương thức thống nhất những biểu ghi (record) ở dạng máy tính đọc được và cho phép tìm kiếm theo chế độ đối thoại. Nhiều hệ thống tìm tin trực tuyến trên thế giới dùng thuật ngữ "file" để chỉ một cơ sở dữ liệu nào đó trong hệ thống.

- **Biểu ghi:** Là tập hợp dữ liệu có cấu trúc được tạo thành bởi một số xác định các thành phần gọi là trường (field). Biểu ghi của cơ sở dữ liệu thư mục là dữ liệu được cấu trúc mà ở đó mỗi thuộc tính mô tả từng đối tượng quản lý của cơ sở dữ liệu được quy định bằng một trường.

- **Trường:** Là thành phần rất cơ bản của biểu ghi. Trường thường được xác định bằng nhãn trường (field tag). Đó là một dãy ký tự nối tiếp nhau (là ký tự chữ cái hay một cụm chữ số) để xác định mô tả một tiêu thức hay một thuộc tính của đối tượng.

Ví dụ, các trường của biểu ghi thư mục theo khổ mẫu MARC 21 được xác định bằng con số như 001, 003, 005, 100, 110, 200, 245, 600,

v.v... Các trường của cơ sở dữ liệu xây dựng bằng chương trình CDS/ISIS cũng được thể hiện bằng các con số.

Nội dung của mỗi trường (nói cách khác là của mỗi thuộc tính) là giá trị trường. Một biểu ghi của cơ sở dữ liệu là tập hợp những giá trị của các trường mô tả một đối tượng.

Ví dụ: Biểu ghi thư mục của một cơ sở dữ liệu về sách có thể có các trường sau:

- ✓ Tiêu đề tác giả.
- ✓ Nhan đề và thông tin trách nhiệm.
- ✓ Lần xuất bản.
- ✓ Địa chỉ xuất bản (nơi xuất bản, nhà xuất bản và năm xuất bản).
- ✓ Thông tin về đặc trưng vật lý (số trang, khổ cỡ,...).
- ✓ Số đăng ký cá biệt (số đăng ký tài sản).
- ✓ Ký hiệu phân loại.
- ✓ Ký hiệu xếp giá (Ký hiệu phân loại, ký hiệu tác giả, năm xuất bản).
- ✓ Từ khoá.
- ✓ Tóm tắt.

Mỗi trường dữ liệu đó sẽ được xác định bởi một nhãn trường. Trong nhiều cơ sở dữ liệu thư mục, trường dữ liệu còn có thể được chia thành các trường con (subfields). Một trường còn thể có nhiều trường con.

Tập hợp các biểu ghi thư mục như vậy được tổ chức thành cơ sở dữ liệu thư mục. Thông tin được lưu trữ trong các cơ sở dữ liệu là dạng thông tin số hoá. Bảng 3 trình bày đẳng cấp của các thành phần dữ liệu của ngân hàng dữ liệu.

Bảng 3. Thí dụ về đẳng cấp của các thành phần dữ liệu

Thành phần dữ liệu	Ví dụ	Giá trị
Bit	1 hoặc 0	
Byte	01001100	
Trường con	Nơi xuất bản	^aHà Nội
Trường	Địa chỉ xuất bản	^aHà Nội^bNXB Khoa học và Kỹ Thuật^c1990
Biểu ghi	- Nhan đề - Tác giả cá nhân - - Địa chỉ xuất bản - Đặc trưng số lượng	
Cơ sở dữ liệu	CSDLBOOK của Mạng VISTA có 90.000 biểu ghi	
Ngân hàng dữ liệu	VISTA	Gồm nhiều CSDL như BOOK, TAPCHI, STD, SCITEC

- **Cơ sở dữ liệu (CSDL):** Được tạo ra, quản lý và khai thác bằng phần mềm máy tính, thường gọi là hệ quản trị CSDL. Hệ quản trị CSDL cho phép người sử dụng có thể thực hiện những chức năng cơ bản sau đây trong việc quản lý dữ liệu: mô tả dữ liệu, cập nhật và bảo trì dữ liệu, tìm kiếm dữ liệu lưu trữ.

Đối tượng quản lý thông tin trong CSDL rất đa dạng. Đó có thể là các thông tin về những bài tạp chí, sách, tuyển tập, luận án tiến sỹ, sáng chế, tiêu chuẩn, hồ sơ cá nhân, hồ sơ công ty, số liệu thống kê, công thức hoá học, cấu tạo các chất, v.v...

Phân loại cơ sở dữ liệu

Có nhiều cách phân loại cơ sở dữ liệu khác nhau phụ thuộc vào tiêu chí phân loại được chọn. Trong hoạt động thông tin-thư viện người ta có thể phân loại cơ sở dữ liệu theo một số dấu hiệu đặc trưng sau:

- Theo tính chất dữ liệu

Dựa vào tính chất dữ liệu, các cơ sở dữ liệu có thể chia thành hai nhóm:

+ Cơ sở dữ liệu tư liệu: dữ liệu trong các cơ sở dữ liệu đó là tin tức về bản thân tài liệu. Cơ sở dữ liệu tư liệu có thể chia thành:

- Cơ sở dữ liệu thư mục: chứa các thông tin thư mục.
- Cơ sở dữ liệu toàn văn.
- Cơ sở dữ liệu dữ kiện.

CSDL thư mục: Chứa các thông tin mô tả thư mục để nhận dạng về tài liệu, chi dẫn cho người dùng tin về nguồn chứa nội dung đầy đủ của tài liệu, giúp họ có khả năng lựa chọn nhanh chóng tài liệu cho các mục đích khác nhau trong cuộc sống hàng ngày.

Thông thường CSDL thư mục sẽ đưa thêm thông tin như từ khóa, chủ đề, chỉ số phân loại và trong nhiều trường hợp có thêm tóm tắt hoặc chú giải.

CSDL toàn văn chứa không chỉ chứa thông báo về tài liệu mà chứa toàn bộ văn bản, nội dung đầy đủ của tài liệu.

Cơ sở dữ liệu dữ kiện chứa các thông tin dạng số liệu, dữ kiện cụ thể, các tính chất hoặc phương pháp nghiên cứu, quy trình công nghệ,...

- Theo phạm vi bao quát của đề tài:

Theo phạm vi bao quát đề tài có thể chia CSDL thành:

+ **Cơ sở dữ liệu đa ngành:** Nội dung của dữ liệu chứa trong CSDL thuộc nhiều ngành khoa học (ví dụ: cơ sở dữ liệu Pascal (Pháp); cơ sở dữ liệu sách (Thư viện Quốc gia);

+ **Cơ sở dữ liệu chuyên ngành:** Nội dung của các dữ liệu thuộc về một ngành cụ thể (Cơ sở dữ liệu Agrindex (FAO), CSDL^Đ dầu khí. (Trung tâm thông tin dầu khí Việt Nam);

+ **Cơ sở dữ liệu theo chủ đề:** Nội dung của các dữ liệu phản ánh theo một chủ đề xác định; CSDL về dân số (Trung tâm thông tin Tổng cục thống kê); CSDL tuyển sinh (Bộ Giáo dục và Đào tạo).

- Theo loại hình tài liệu.

Một số cơ sở dữ liệu này định hướng vào một loại hình tài liệu cụ thể và thường có nội dung đa ngành.

+ CSDL sách

+ CSDL tạp chí

+ CSDL báo cáo

+ CSDL sáng chế

+ CSDL tiêu chuẩn,...

Ví dụ về một số cơ sở dữ liệu theo loại hình tài liệu ở Việt Nam:

✓ Cơ sở dữ liệu sách (Thư viện Quốc gia Việt Nam).

✓ Cơ sở dữ liệu báo (Thư viện Quốc gia Việt Nam).

- ✓ Cơ sở dữ liệu LA (Thư viện Quốc gia Việt Nam).
- ✓ Cơ sở dữ liệu KQNC chuyên về báo cáo kết quả nghiên cứu (Trung tâm Thông tin KH và CN Quốc gia).

- Phân loại theo tính chất của dữ liệu:

Trong hoạt động tìm tin trực tuyến, có một số tác giả, dựa tính chất của dữ liệu, chia CSDL các NHDL thành hai nhóm lớn:

Cơ sở dữ liệu tra cứu (*Reference Databases*).

Cơ sở dữ liệu gốc (*Source Databases*).

+ *Cơ sở dữ liệu tra cứu*: Là những CSDL chứa các mẫu tìm đại diện cho nguồn tin cấp một. Các CSDL tra cứu có mục tiêu chỉ ra nguồn chứa thông tin mà không chứa nội dung đầy đủ của nguồn tin cấp một. CSDL tra cứu có thể được chia thành hai nhóm :

CSDL thư mục (*Bibliographic Databases*).

CSDL chỉ dẫn (*Referral Databases*).

- CSDL chỉ dẫn là nhóm CSDL chỉ dẫn đến nguồn tin không phải tài liệu in (non-printed information sources) đó là những nguồn tin đồng nhất hơn như về con người, cơ quan/tổ chức, doanh nghiệp, đề tài nghiên cứu,... Các CSDL chỉ dẫn là danh bạ tin học hoá (directories), cho phép tìm kiếm tên và/hoặc địa chỉ của một cá nhân hoặc một tổ chức. Cơ sở dữ liệu chỉ dẫn có thể có chứa những trường dữ liệu số (dữ kiện). Mặc dù có thể tìm kiếm theo chủ đề, tìm kiếm theo tên là một trong những yêu cầu cần có của các cơ sở dữ liệu thuộc lớp này.

+ *Cơ sở dữ liệu gốc*: Là những CSDL liệu chứa thông tin cấp một (dữ liệu gốc) hoặc những dữ liệu được rút ra từ nguồn tin mà khi có chúng người dùng tin không cần phải tham khảo lại nguồn chứa dữ liệu.

Các biểu ghi trong những cơ sở dữ liệu này không chỉ dẫn người sử dụng tới những cuốn sách hay những bài báo mà cung cấp cho họ tức thì những thông tin đầy đủ.

Các CSDL gốc có thể chia thành các nhóm:

- Cơ sở dữ liệu Số (Numeric databases).
- Cơ sở dữ liệu văn bản-số (Text-Numeric databases).
- Cơ sở dữ liệu tính chất (Properties databases).
- Cơ sở dữ liệu toàn văn (Fulltext databases).

Cơ sở dữ liệu số (Numeric databases), bao gồm thông tin là các số liệu được trình bày dưới dạng bảng hoặc các chuỗi số liệu theo thời gian (time series).

Ví dụ: Những số liệu về dân số, số liệu thống kê qua một giai đoạn, về sản xuất, nhập khẩu, xuất khẩu, về thị trường chứng khoán,...

Cơ sở dữ liệu văn bản - số (Text-Numeric databases) là những CSDL chứa một tổ hợp văn bản và thông tin dữ kiện. Nhiều cơ sở dữ liệu cung cấp thông tin về công ty là thuộc loại này: mô tả công ty, các dữ kiện về tình hình kinh doanh của công ty (như doanh thu, lợi nhuận, vốn,...).

Các cơ sở dữ liệu tính chất (Properties databases) chứa thông tin về các tính chất riêng về vật lý, hóa học và cơ học của các dung môi và các hợp chất.

Ví dụ:

✓ CSDL liệu Heibron (từ điển về các hợp chất vô cơ và những từ điển khác,

✓ CSDL Sách tra cứu về hóa vô cơ của Beilstein, Martindale Extra Pharmacopoeia,

- ✓ Bách khoa thư về công nghệ hóa học Kir-Othmer,
- ✓ Hệ thống tìm tin cấu trúc hóa học DARC.

CSDL cung cấp cho cán bộ quản lý, các kỹ sư, cán bộ kỹ thuật và cơ sở sản xuất những thông tin cụ thể, đảm bảo độ tin cậy, giúp họ trong xuất nghiên cứu, kinh doanh có hiệu quả nhất.

2.2.3.4. Ngân hàng dữ liệu (NHDL)

Ngân hàng dữ liệu là một khái niệm tương đối. Ngân hàng dữ liệu có thể được định nghĩa là "một hệ thống thông tin tự động hoá, khai thác trực tiếp, được cấu trúc thành cơ sở dữ liệu bao quát một lĩnh vực nào đó của tri thức".

Ngân hàng dữ liệu có các thành phần sau:

- Một tập hợp các cơ sở dữ liệu khác nhau, trong đó có thể gồm các cơ sở dữ liệu do nhân hàng tự xây dựng lấy hoặc bổ sung từ các cơ quan/tổ chức bên ngoài.
- Tổ hợp các phương tiện xây dựng, bảo trì và khai thác cơ sở dữ liệu, bao gồm:
 - + Máy tính điện tử và các thiết bị ngoại vi;
 - + Các thiết bị truyền và nhận dữ liệu từ xa;
 - + Các thiết bị sao chụp;
 - + Các phần mềm quản trị các CSDL;
 - + Các phần mềm giao tiếp giữa người và máy;
 - + Các thủ tục và phương pháp truyền và truy cập thông tin.
- Nhân lực làm việc với hệ thống: các nhân viên quản trị, vận hành và khai thác CSDL...

Ngân hàng dữ liệu có thể có hàng trăm CSDL, phong phú về diện bao quát đề tài, về loại hình dữ liệu, phục vụ được các loại khách hàng với nhiều yêu cầu đa dạng khác nhau.

Trong kinh doanh thông tin, người ta quan niệm các cơ quan cung cấp dịch vụ trực tuyến tới nhiều CSDL là NHDL. Đó là các tổ chức bán dịch vụ truy cập trực tuyến. Một ngân hàng dữ liệu có thể được gọi là nhà bán hàng trực tuyến (vendor), là một tổ chức cung cấp cho khách hàng thông tin hoặc cho các chuyên gia thông tin giúp khách hàng truy cập trực tuyến tới các CSDL dưới dạng máy có thể đọc được. Các NHDL có thể thu thập CSDL từ các cơ quan sản xuất CSDL, hoặc trong một số trường hợp, tự sản xuất ra các CSDL mà ngân hàng đó tin rằng sẽ được khách hàng sử dụng.

NHDL có các hệ thống phần mềm cho phép tìm tin trong những CSDL thuộc ngân hàng dữ liệu. Ngân hàng này cũng cho thuê hoặc bán và bảo trì các máy tính lớn và các thiết bị phụ trợ, viết tài liệu mô tả và hướng dẫn khai thác các cơ sở dữ liệu của họ.

Hầu hết các NHDL là các tổ chức thương mại. Tất cả đều thực hiện các dịch vụ tìm kiếm thông tin. Thông thường giá cả truy cập thường được tính theo thời gian kết nối từ máy của người tìm tin tới máy chủ và khối lượng thông tin khai thác được.

Ví dụ:

Các dịch vụ của **DIALOG**, **STN International** đều tính chi phí khai thác gồm hai phần:

- ✓ Thời gian truy cập hệ thống để khai thác tài nguyên thông tin và dịch vụ kỹ thuật của hệ thống.
- ✓ Chi phí dựa trên số lượng biểu ghi tìm được, được xem và tải về.

Số lượng các NHDL trên thế giới tăng lên nhanh chóng. Cuadra (1978) xác định được năm 1977 thế giới có 15 NHDL tầm cỡ quốc tế, trong đó có BRS, DIALOG, Mead Corporation, OCLC và SDC.

Đến năm 1991 thế giới đã có 731 NHDL; trong số đó một phần đáng kể là các hãng cung cấp truy cập trực tuyến tới nguồn các CSDL.

Tuy nhiên, cũng có các tổ chức không kinh doanh kiếm lời, có nhiều CSDL cho phép truy cập CSDL là các ngân hàng dữ liệu.

Với sự phát triển của Internet, công nghệ thông tin, viễn thông, đặc biệt là xuất bản điện tử, ngày càng xuất hiện nhiều nhà xuất bản cung cấp khả năng truy cập đến các tạp chí do họ xuất bản dưới dạng điện tử. Xuất hiện một loại hình tổ chức dịch vụ mới gọi là "*các tổ hợp thông tin*" (Aggregators). Tổ hợp thông tin là một tổ chức thu thập, tổ chức và cung cấp dịch vụ truy cập đến tạp chí điện tử do họ phát hành hoặc mua trên mạng Internet. Các tổ hợp này cung cấp cả khả năng truy cập thông tin thư mục thông qua các cơ sở dữ liệu. Một số tổ hợp tin lớn như ScienceDirect do Nhà xuất bản Elsevier xây dựng hoặc EBSCOHost ở Hoa Kỳ cung cấp khả năng truy cập tập trung đến hàng chục nghìn tạp chí điện tử toàn văn.

2.3. Lệnh tìm

Lệnh tìm là nội dung yêu cầu thông tin được diễn đạt bằng những thuật ngữ của ngôn ngữ tìm tin. Như đã nghiên cứu ở trên, ngôn ngữ tìm tin có thể là: Chỉ số phân loại, đề mục chủ đề hoặc từ (cụm từ) được sử dụng làm từ khoá, từ chuẩn như: thuật ngữ khoa học, địa danh, tên nhân vật, tên niên hiệu chỉ thời gian... có hoặc không có các quan hệ được thiết lập giữa chúng với nhau (sử dụng các toán tử logic OR, AND, NOT).

Khi lập lệnh tìm cần xem xét kỹ, hệ thống sử dụng loại ngôn ngữ tìm tin nào để mô tả nội dung tài liệu, trên cơ sở đó lệnh tìm cũng được thể hiện bằng ngôn ngữ đó.

Ví dụ: Thư viện đó sử dụng bảng phân loại nào? UDC, DDC hay BBK; Bảng từ khoá của TVQG hay Từ điển từ khoá khoa học - công nghệ của Trung tâm Thông tin KH&CN Quốc gia? Trong quá trình hoạt động các thư viện, cơ quan thông tin sử dụng song song nhiều loại ngôn ngữ tìm tin khác nhau để mô tả nội dung tài liệu, tổ chức các hình thức lưu trữ thông tin đa dạng và giúp cho việc tra cứu thông tin đầy đủ, nhanh chóng, thuận tiện.

Nếu sử dụng ngôn ngữ tìm tin bằng từ khoá để tra cứu tự động hoá thì cần phải xây dựng phương trình tìm.

Phương trình tìm gồm có từ khóa kết hợp với các toán tử tìm. Việc áp dụng cú pháp của ngôn ngữ tìm tin trong việc xây dựng lệnh tìm, biểu thức trong tìm tin tự động hoá được trình bày chi tiết tại Chương IV và V.

2.4. Con người

Đó là những người có tác động trực tiếp tới hệ thống, gồm những người có liên quan trong hệ thống: người tìm tin và người dùng tin.

- Người tìm tin/ Cố vấn bạn đọc:

Tra cứu là công việc trọng tâm của tất cả các thư viện và cơ quan thông tin. Đa phần người dùng tin đến thư viện với mục đích rõ ràng và với nhu cầu tin xác định, họ biết mình cần thông tin gì? phải tra cứu ở đâu? phương pháp tra cứu như thế nào? Tuy nhiên, cũng không ít người dùng tin không biết sẽ bắt đầu từ đâu. Như vậy là họ sẽ phải cần tới sự giúp đỡ của cán bộ thư viện - người tìm tin. Người tìm tin có thể là thủ thư nói riêng hoặc cán bộ thư viện nói

chung. Người tìm tin có thể trực tiếp trả lời câu hỏi cho người dùng tin hoặc giúp họ tra cứu thông tin cần thiết. Người tìm tin giúp người dùng tin nắm được và sử dụng nguồn lực thông tin có trong thư viện. Để giúp người dùng tin lựa chọn thông tin phù hợp với nhu cầu của họ, do đó người tìm tin cần:

+ *Kỹ năng chung:*

- Có kiến thức hiểu biết chung tốt,
- Có khả năng phán đoán và phân tích vấn đề,
- Lịch thiệp,
- Thân mật,

+ *Kỹ năng giao tiếp:*

- Có kỹ năng nghe và đặt câu hỏi,
- Có khả năng chỉ dẫn, giải thích vấn đề,
- Có khả năng giúp người dùng tin không phân biệt lứa tuổi, tính cách, sở thích khác nhau của họ.

+ *Kỹ năng nghiệp vụ thông tin thư viện:*

- Nắm vững nguồn tin/ vốn tài liệu, am hiểu nguồn tin có trong cơ quan: nguồn tin cấp một, nguồn tin cấp hai, ba và những nguồn tin ở ngoài, đặc biệt là nguồn tin trên mạng và những đặc thù riêng của kho tài liệu tra cứu,

- Có hiểu biết nghiệp vụ về thông tin - thư viện học: chính sách bổ sung và những quy tắc, yêu cầu, quy trình hoạt động của cơ quan thông tin thư viện và kiến thức liên quan tới lĩnh vực thông tin cần tra cứu.

- Có kỹ năng tra cứu thành thực, nắm được chiến lược tìm tin, cách trình bày và biết sử dụng những phương tiện kỹ thuật nhằm đáp ứng yêu cầu thông tin.

- Có khả năng quyết định những yêu cầu tin nào có thể giúp NDT tham khảo hoặc tra cứu, những yêu cầu cần chi dẫn họ tới thư viện hoặc cơ quan thông tin khác phù hợp,
- Có những hiểu biết về nguyên tắc và các dịch vụ khách hàng,
- Có sự ham thích, say sưa với công tác tra cứu.

Để đáp ứng với những yêu cầu trên, người tìm tin không bao giờ thoả mãn với kiến thức mình có. Nếu không thể trả lời được câu hỏi của người dùng tin, cần hướng dẫn họ tới những người, những nơi có thể giúp họ tra cứu thông tin cần thiết. Cần phải thường xuyên cập nhật thông tin cả những vấn đề về nghiệp vụ cũng như những vấn đề chuyên môn sâu (nếu làm việc ở các cơ quan thông tin - thư viện chuyên ngành) và những vấn đề mang tính thời sự - chính trị được đăng tải trên các phương tiện thông tin đại chúng như: báo, tạp chí, đài truyền hình, đài phát thanh Trung ương và địa phương. Trong thực tế không ít trường hợp những yêu cầu tin xuất phát từ đó.

Người dùng tin (NDT)

Người dùng tin - là yếu tố cơ bản của mọi hệ thống thông tin. Người dùng tin là đối tượng phục vụ của công tác thông tin - thư viện, họ vừa là khách hàng của các dịch vụ thông tin, đồng thời họ cũng có thể là người sản sinh ra thông tin mới.

Người dùng tin giữ vai trò quan trọng trong các hệ thống tìm tin. Họ là yếu tố tương tác hai chiều với các đơn vị thông tin, là cơ sở để định hướng các hoạt động của một đơn vị thông tin: họ tham gia vào hầu hết các công đoạn của dây chuyền hoạt động thông tin thư viện; Họ biết những nguồn thông tin và có thể thông báo và đánh giá các nguồn thông tin đó; NDT có khả năng giúp đỡ trong việc lựa chọn,

bổ sung vốn tài liệu (chính sách bổ sung phụ thuộc vào nhu cầu tin của người dùng tin); họ có thể tham gia xây dựng ngôn ngữ tìm tin, xác định cấu trúc các bộ phiếu, xử lý thông tin; hình thành chiến lược tra cứu và đánh giá kết quả tìm tin.

Tóm lại, để đáp ứng nhu cầu thông tin trong kỷ nguyên mới với chủ trương phát huy nội lực trong phát triển kinh tế xã hội, việc tìm tin và cung cấp thông tin từ các nguồn tin trong nước, đặc biệt là nguồn tin "xám" có ý nghĩa quan trọng vì nó giúp cho cán bộ khoa học nghiên cứu không bị trùng lặp, mặt khác vấn đề tra cứu thông tin mới trên các mạng, đặc biệt trên mạng Internet phục vụ cho công tác nghiên cứu - triển khai, phát triển kinh tế - xã hội, hội nhập kinh tế quốc tế giữ vai trò quan trọng trong hoạt động thông tin thư viện.

Trên cơ sở nguồn thông tin hiện có, tổ chức tốt bộ máy tra cứu đa dạng, phong phú, linh hoạt với các trang thiết bị mới, bộ tra cứu thông tin của cơ quan thông tin - thư viện sẽ đáp ứng nhu cầu tin trong công cuộc đổi mới.

Chương 2

Chiến lược tra cứu thông tin

Trong đây chuyên hoạt động thông tin tư liệu, tra cứu thông tin (tìm tin) là công đoạn quan trọng. Để thực hiện tra cứu có hiệu quả, đáp ứng yêu cầu tin phong phú, đa dạng của người dùng tin cần có chiến lược tra cứu khoa học, logic, có bộ máy tra cứu đầy đủ và người tìm tin có kinh nghiệm, linh hoạt.

Để tiến hành tra cứu thông tin cần xác định các yếu tố của chiến lược tìm tin. Các yếu tố đó gồm:

Lệnh tìm:

Lệnh tìm là nội dung yêu cầu thông tin được diễn đạt bằng những thuật ngữ của ngôn ngữ tìm tin. Lệnh tìm có thể được thể hiện bằng:

- Các chỉ số phân loại.

Theo Bảng phân loại dùng cho các thư viện khoa học tổng hợp (2002):

3. Các khoa học xã hội - Chính trị

✓ Tìm các tài liệu có đề cập tới Các khoa học xã hội - Chính trị.

▪ **Lệnh tìm KHPL: 3**

55. Địa lý tự nhiên. Vật lý địa cầu. Địa chất học

✓ Tìm các tài liệu có đề cập tới Địa lý tự nhiên. Vật lý địa cầu.
Địa chất học.

▪ **Lệnh tìm** KHPL: 55

6C7.7 Các chất sơn và các chất phẩm màu.

✓ Tìm các tài liệu có đề cập tới Các chất sơn và các chất phẩm màu

▪ **Lệnh tìm** KHPL: 6C7.7

- Các đề mục chủ đề:

Thực phẩm

✓ Tìm các tài liệu có đề cập tới Thực phẩm.

▪ **Lệnh tìm:CD** *Thực phẩm*

Vệ sinh môi trường.

Tìm các tài liệu có đề cập tới Vệ sinh môi trường.

▪ **Lệnh tìm: CD** *Vệ sinh môi trường*

Thơ trữ tình.

✓ Tìm các tài liệu có đề cập tới Thơ trữ tình.

▪ **Lệnh tìm:CD** *Thơ trữ tình*

- Từ khoá, từ chuẩn (TK, TC):

TK thuật ngữ khoa học: sắc xuất toán học, truyền hình, đèn hiệu,...

✓ Tìm các tài liệu có đề cập tới sắc xuất toán học; truyền hình;
đèn hiệu;...

▪ **Lệnh tìm TK:** *Sắc xuất toán học; truyền hình; đèn hiệu,...*

TK tên người: Nguyễn Bình Khiêm, Đinh Gia Khánh, Phan Minh,

✓ Tìm các tài liệu có đề cập tới Nguyễn Bình Khiêm, Đinh Gia
Khánh, Phan Minh,

▪ **Lệnh tìm TK:** *Nguyễn Bình Khiêm, Đinh Gia Khánh, Phan
Minh.*

TK tên địa lý: Tây Bắc, Hà Nam, Cù Lao Chàm,...

✓ Tìm các tài liệu có đề cập tới Tây Bắc, Hà Nam, Cù Lao Chàm,...

▪ **Lệnh tìm TK**: *Tây Bắc, Hà Nam, Cù Lao Chàm,*

TK tên niên hiệu chỉ thời gian: Thế kỷ XX, Cách mạng tháng Tám, Kỷ nguyên thông tin.

✓ Tìm các tài liệu có đề cập tới Thế kỷ XX, Cách mạng tháng Tám, Kỷ nguyên thông tin.

▪ **Lệnh tìm TK**: *Thế kỷ XX, Cách mạng tháng Tám, Kỷ nguyên thông tin.*

Có (hay không có) các quan hệ được thiết lập giữa chúng với nhau.

Mẫu tìm:

Mẫu tìm là nội dung, ý nghĩa của tài liệu được diễn đạt bằng những thuật ngữ của ngôn ngữ tư liệu (Để mục chủ đề, kí hiệu phân loại, từ khoá). Mẫu tìm chỉ thể hiện nội dung, ý nghĩa cơ bản của tài liệu mà không phải toàn bộ thông tin chứa trong tài liệu đó. Nội dung này phù hợp với một tài liệu nào đó và dùng nó trong tập hợp nhiều các tài liệu khác.

- **Ký hiệu phân loại**

Theo Bảng phân loại dùng cho các thư viện khoa học tổng hợp (2002):

+ Các tài liệu phản ánh về lĩnh vực khoa học xã hội - Chính trị có KHPL: 3.

Mẫu tìm: 3

+ Các tài liệu phản ánh về lĩnh vực Địa lý tự nhiên. Vật lý địa cầu. Địa chất học có KHPL: 55

Mẫu tìm: 55

+ Các tài liệu phản ánh về lĩnh vực Các chất sơn và các chất phẩm màu có KHPL: 6C7.7.

Mẫu tìm: 6C7.7

+ Các tài liệu có nội dung phản ánh về hoá lí, lí hoá, hoá keo đều được định KHPL: 541.

Mẫu tìm: 541

Tra tìm trong Mục lục phân loại của Hệ thống mục lục thư viện.

- Để mục chủ đề:

+ Các tài liệu có nội dung phản ánh về Thực phẩm, Vệ sinh môi trường, đều có chủ đề “Thực phẩm”, “Vệ sinh môi trường”.

Mẫu tìm CD: *Thực phẩm, Vệ sinh môi trường*

+ Các tài liệu có nội dung phản ánh về Thơ trữ tình, đều có chủ đề “Thơ trữ tình”.

Mẫu tìm CD: *Thơ trữ tình*

Tra tìm trong Mục lục chủ đề của Hệ thống mục lục thư viện

- Từ khoá:

+ Các tài liệu có nội dung phản ánh về toán học sắc xuất, đều có từ khoá “Truyền hình”, “Toán học sắc xuất”.

Mẫu tìm TK: *Truyền hình, toán học sắc xuất.*

+ Các tài liệu có nội dung phản ánh về Nguyễn Bình Khiêm, Đinh Gia Khánh đều có từ khoá chi tên riêng “Nguyễn Bình Khiêm”, “Đinh Gia Khánh”.

Mẫu tìm TK: *Nguyễn Bình Khiêm, Đinh Gia Khánh.*

+ Các tài liệu có nội dung phản ánh về Tên địa lý: Tây Bắc, Hà Nam đều có các từ khoá chi tên địa danh “Tây Bắc”, “Hà Nam”.

Mẫu tìm TK địa danh: *Tây Bắc, Hà Nam.*

+ Các tài liệu có nội dung phản ánh về: Thế kỷ XX, Cách mạng tháng Tám, đều có các từ khóa chi tên niên hiệu chi thời gian: "Thế kỷ XX", "Cách mạng tháng Tám".

Mẫu tìm TK chi tên niên hiệu chi thời gian *Thế kỷ XX, Cách mạng tháng Tám...*

Tra tìm trong *Cơ sở dữ liệu và ấn phẩm thông tin.*

Lưu ý: Khi tra cứu tin theo từ khóa bằng phương pháp tự động hóa phải xây dựng *Biểu thức tìm* (TK + OR/ AND/ NOT).

Mảng tìm:

Mảng tìm là những hình thức lưu trữ thông tin phản ánh nguồn lực của cơ quan thông tin thư viện. Mảng tìm chính là các thành phần của bộ máy tra cứu.

- **Bộ máy tra cứu truyền thống:** Hệ thống mục lục phiếu, các hộp phiếu tra cứu, tài liệu tra cứu dạng giấy, ...

- **Bộ máy tra cứu bán tự động:** Phiếu lỗ mép, phiếu lỗ soi, ...

- **Bộ máy tra cứu tự động hoá:**, các CSDL TM, CSDL DK, mục lục truy nhập công cộng trực tuyến (OPAC), Tài liệu tra cứu điện tử, NHDL, mạng thông tin ...

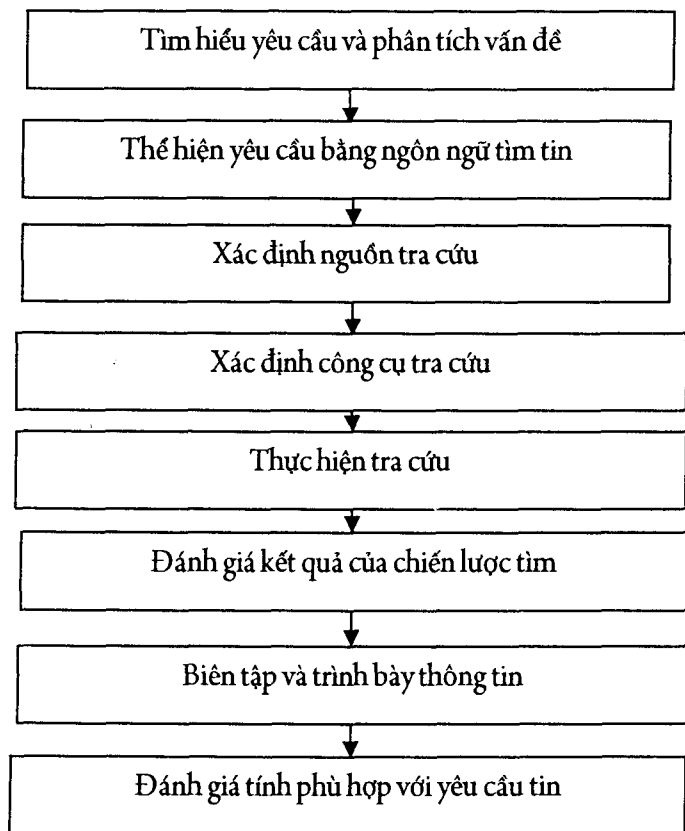
Người tìm tin.

Con người là yếu tố không thể thiếu được trong hoạt động thông tin tư liệu nói chung, cũng như quá trình tìm tin nói riêng. Đó chính là những người trực tiếp tiến hành tra cứu tài liệu/ thông tin cần thiết. Họ có thể là thủ thư/ cán bộ thư viện và cũng có thể là người dùng tin. Trong quá trình tra cứu, nếu lệnh tìm và mẫu tìm trùng hợp nhau,

thông tin có trong màng tìm, có nghĩa là quá trình tra cứu tin đạt kết quả, yêu cầu của người dùng tin được đáp ứng. Để tiến hành tra cứu cần tiến hành các bước một cách hợp lý và lôgic.

1. CÁC BƯỚC CỦA QUÁ TRÌNH TRA CỨU/TÌM TIN

Các bước của quá trình tra cứu có thể được trình bày như trong hình 2.



Hình 2. Sơ đồ quá trình tìm tin

1.1. Tìm hiểu yêu cầu tin và phân tích vấn đề

Cùng với sự phát triển mạnh mẽ của khoa học công nghệ các nhu cầu xã hội ngày càng không ngừng gia tăng. Các nhu cầu hiện nay không chỉ bó hẹp trong phạm vi về ăn, mặc, ở, đi lại... mà còn có nhiều nhu cầu khác, trong đó có nhu cầu về thông tin (Nhu cầu tin).

Nhu cầu tin của người dùng tin là cơ sở quan trọng để tiến hành các hoạt động thông tin. Nhu cầu tin xuất phát từ nhu cầu nhận thức: muốn được hiểu biết và khám phá về các sự vật, hiện tượng, quy luật,... Một khi con người có nhu cầu nhận thức, tất yếu họ sẽ có nhu cầu thông tin ổn định.

Nhu cầu tin là tính chất của một đối tượng cá nhân, tập thể hoặc một hệ thống nào đó thể hiện sự cần thiết nhận thông tin phù hợp với hành vi hay công việc mà đối tượng đó đang thực hiện. Nhu cầu tin vừa mang tính chủ quan và vừa mang tính khách quan.

Nhu cầu tin mang tính chủ quan vì nó gắn với người dùng tin nhất định. Nhu cầu tin phụ thuộc vào kiến thức, trình độ của từng người dùng tin, và cũng mang tính khách quan vì nó phát sinh từ hoạt động thực tiễn của con người và xã hội mà họ sinh sống.

Nhu cầu tin có nhiều loại, trong quá trình tìm hiểu nhu cầu tin của người dùng tin, cần phân biệt:

Nhu cầu tin tuyệt đối: đó là sự cần thiết nhận thông tin một cách đầy đủ, trực tiếp, chính xác, kịp thời để giải quyết công việc.

Nhu cầu tin thực tế: là sự cần thiết, đòi hỏi của người dùng tin về thông tin và được thoả mãn bằng những thông tin hiện có trong cơ quan thông tin thư viện đó. Nhu cầu tin thường tồn tại trong thời gian

tương đối dài. Người dùng tin đến cơ quan thông tin thư viện và thể hiện nhu cầu tin bằng những yêu cầu tin cụ thể. Để thể hiện trọn vẹn một nhu cầu tin, NDT phải thể hiện hàng loạt các yêu cầu tin, đồng thời có sự liên hệ, trao đổi hiệu chỉnh thường xuyên giữa cơ quan thông tin thư viện và người dùng tin.

Yêu cầu tin là sự thể hiện một khía cạnh của nhu cầu tin được diễn đạt bằng ngôn ngữ tự nhiên, nó được cụ thể hoá bằng các câu hỏi. Để trả lời được câu hỏi một cách đầy đủ, cán bộ tra cứu cần thiết phải phân tích nhằm:

- Xác định câu hỏi để làm rõ nội dung của vấn đề:

+ *Câu hỏi đó thuộc lĩnh vực nào?*

Thông thường, các câu hỏi của người dùng tin đưa ra là những lĩnh vực, vấn đề thuộc diện phục vụ của cơ quan thông tin - thư viện. Nếu gặp câu hỏi với những vấn đề mới, người tìm tin có thể tham khảo vấn đề đó trong các từ điển chuyên ngành, từ điển thuật ngữ, các loại bách khoa thư có trong thư viện hoặc từ điển từ chuẩn.

+ *Chọn các thuật ngữ, lĩnh vực khoa học* liên quan tới câu hỏi. Những thuật ngữ này cần chọn trong bảng từ khoá, từ chuẩn, bảng đề mục chủ đề, hoặc bảng phân loại mà cơ quan thông tin - thư viện sử dụng để mô tả nội dung tài liệu, giúp cho việc tìm tin nhanh chóng, chính xác và không bị mất tin.

PHIẾU YÊU CẦU TIN

Họ và tên:.....

Chức vụ:.....

Trình độ: ă Tiến sỹ ă Thạc sỹ ă Đại học ă Sinh viên

Địa chỉ liên lạc:.....

.....

Điện thoại: Fax: Email:

Yêu cầu tin:

Ngôn ngữ: ă Tất cả ă Anh ă Nga ă Pháp

Khác:..... Độ hồi có: từ năm.....

Gợi ý từ khoá:.....

Những bài báo đã biết:

1.

2.

Thời gian cần kết quả:

Hình 3. Ví dụ về biểu mẫu nhận yêu cầu tìm tin

+ *Xác định mức độ chi tiết/ độ rộng* của vấn đề. Điều này phụ thuộc vào đối tượng người dùng tin: Ai sẽ sử dụng thông tin được cung cấp? Trình độ học vấn? Trình độ chuyên môn? Trình độ ngoại ngữ? Sở thích cá nhân? Họ nghiên cứu vấn đề nhằm mục đích gì? Để phục vụ

cho công tác nghiên cứu? học tập? ứng dụng - triển khai hay nâng cao dân trí?

- **Xác định thời gian xuất bản của tài liệu.** Người dùng tin cần tài liệu/thông tin xuất bản khi nào? Từ đó xác định họ cần thông tin cập nhật (thông tin/ tài liệu mới) , thông tin hồi cố (thông tin/ tài liệu đã xuất bản đã lâu) hay thông tin dự báo (thông tin/ tài liệu sẽ xuất bản trong thời gian tới)?

- **Xác định thời gian cung cấp thông tin:** Người dùng tin cần tin tức thời, định kỳ hay vào thời điểm nhất định nào đó?.

- **Xác định ngôn ngữ của tài liệu:** Người dùng tin có khả năng sử dụng được những tài liệu bằng tiếng Nga, tiếng Anh, tiếng Trung?...

- **Xác định hình thức xuất bản của tài liệu:** Người dùng tin cần: Sách, báo, tạp chí, catalo công nghiệp (mục lục công nghiệp), tranh ảnh, bản đồ; Dạng tài liệu (giấy, dạng điện tử), phim ảnh, băng hình hay đĩa nhạc?....

- **Xác định hình thức cung cấp thông tin** (sản phẩm đầu ra): Có thể tài liệu gốc (bản sao); chương hoặc sơ đồ nào đó trong một/ tập tài liệu cụ thể, danh mục tài liệu, tóm tắt tài liệu, các số liệu, biểu, bảng, trích lục bản đồ ?...

Để tra cứu được thông tin chính xác và phù hợp, cần thiết phải trao đổi với người dùng tin để làm rõ các vấn đề liên quan tới câu hỏi, có kết luận sơ bộ: vấn đề đó có nằm trong diện bao quát đề tài của cơ quan thông tin thư viện mình hay không? Nếu không, cần chi cho người dùng tin đến nơi khác phù hợp với yêu cầu của họ. Nếu thuộc diện phục vụ, sẽ tiếp tục các bước tiếp theo nhằm lựa chọn tài liệu cho thích hợp với từng đối tượng cụ thể.

Để hỗ trợ cho tìm hiểu yêu cầu tin và phân tích vấn, người cán bộ thực hiện tra cứu cần sử dụng các phiếu yêu tin theo mẫu và yêu cầu người dùng tin điền chi tiết vào phiếu. Một ví dụ về phiếu yêu cầu tin được trình bày trong hình 3.

1.2. Thể hiện yêu cầu bằng ngôn ngữ tìm tin

Các yêu cầu tin/câu hỏi của người dùng tin được thể hiện bằng ngôn ngữ tự nhiên. Tuy nhiên với ngôn ngữ đó không phải lúc nào chúng ta cũng tra cứu được tài liệu/ thông tin họ cần, đặc biệt là khi tìm tin theo nội dung. Như vậy để tra cứu thông tin theo nội dung cần chuyển các yếu tố câu hỏi sang các yếu tố ngôn ngữ tìm tin. Các yếu tố này có thể được thể hiện bằng một/ một số loại ngôn ngữ tìm tin:

- **Ký hiệu phân loại** (Ngôn ngữ tìm tin theo hệ thống đẳng cấp),
- **Chủ đề** (Ngôn ngữ tìm tin theo đề mục chủ đề),
- **Biểu thức tìm** (Ngôn ngữ tìm tin theo từ khoá : Kết hợp các từ khoá với toán tử Boole). Ngôn ngữ đó phải tương thích với ngôn ngữ tư liệu được sử dụng trong quá trình xử lý nội dung của cơ quan thông tin thư viện và nơi quá trình tra cứu được tiến hành.

1.3. Xác định nguồn tra cứu

Nguồn tra cứu trong các thư viện và cơ quan thông tin rất đa dạng và phong phú. Dựa vào câu hỏi để xác định:

- Thông tin cần tra cứu thường có ở đâu? Có trong cơ quan thông tin - thư viện của mình, các thư viện khác, cơ quan lưu trữ, phòng tư liệu, cơ quan thông tấn báo chí, phòng thương mại ...

- Hình thức và loại hình xuất bản nào phù hợp? Báo, tạp chí, sách, luận án, tài liệu chuyên dạng; Dưới dạng giấy, vi hình, vi phiếu, tài liệu điện tử?

Hiện nay thực hiện việc tra cứu không chỉ giới hạn ở trong một thư viện cụ thể nào đó, mà không ít trường hợp phải khai thác tài liệu/thông tin từ bên ngoài qua các mạng thông tin trong nước và quốc tế để có được những thông tin mới, đầy đủ đáp ứng yêu cầu.

1.4. Lựa chọn công cụ tra cứu

Ở các cơ quan thông tin thư viện, công cụ tra cứu rất phong phú, đa dạng, phản ánh được nhiều khía cạnh về nội dung, hình thức, ngôn ngữ cũng như bao quát các loại hình của tài liệu. Công cụ tra cứu trong cơ quan thông tin thư viện thường được tổ chức thành:

- Công cụ tra cứu truyền thống: Hệ thống mục lục dạng phiếu, các bộ phiếu tra cứu, kho tài liệu tra cứu,...

- Công cụ tra cứu tự động hoá: ngân hàng dữ liệu, cơ sở dữ liệu, OPAC, tài liệu tra cứu điện tử,...

Với các yêu cầu tin cụ thể đã xác định, cần lựa chọn và sử dụng công cụ tra cứu nào phù hợp với các nguồn thông tin.

Để tra cứu nhanh, đầy đủ cần xác định các nguồn tra cứu chỉ dẫn tới các nguồn tin. Các nguồn tin đó gồm: *Các bảng tra cứu, bảng chỉ dẫn* trong các Bách khoa thư, từ điển bách khoa, ấn phẩm thông tin,...

Đối với những vấn đề có nội dung chuyên ngành hẹp, người dùng tin cần cung cấp thông tin một cách chi tiết cho người tìm tin, khi họ chưa có điều kiện cập nhật hoặc không thật chuyên sâu. Mặt khác, người tìm tin có thể tiếp xúc, gặp gỡ, trao đổi trực tiếp với các

chuyên gia/ cộng tác viên của thư viện trong lĩnh vực chuyên môn đó, vì các cộng tác viên là những người rất am hiểu nguồn lực thông tin của thư viện, thông thạo phương pháp tra cứu, để họ giúp đỡ làm rõ vấn đề hoặc gợi ý nên tra cứu trong CSDL, nguồn tài liệu tra cứu nào... mà cán bộ tra cứu chưa biết hay không có tại thư viện.

1.5. Thực hiện tra cứu

Để thực hiện tra cứu cần hai yếu tố: khoá tìm và khoá truy nhập

- **Khoá tìm:** Là yếu tố phản ánh những đặc tính khác nhau của đối tượng, được sử dụng trong quá trình tra cứu tin và lựa chọn thông tin. Người dùng tin biểu thị chúng trong yêu cầu của mình khi phản ánh các thuộc tính về nội dung và hình thức của thông tin.

- **Khoá truy nhập:** Là các thuật ngữ phản ánh về chủ đề, môn loại của đối tượng: ngành, chuyên ngành khoa học - công nghệ, loại đơn vị sản xuất kinh doanh, loại cơ quan nghiên cứu, triển khai,...; Các thông tin về tác giả; loại đối tượng thông tin: tài liệu in, tài liệu điện tử,...; về vùng địa lý: Tên châu lục, vùng, nước, tên địa danh cụ thể,...; Danh nhân/ Nhân vật: Những người nổi tiếng trong nước và trên thế giới,...; khoảng thời gian có liên quan tới đối tượng mà yêu cầu tin hướng tới, ...

Khi tra cứu tin cần phối hợp các khoá tìm và khoá truy nhập.

+ *Khi tra cứu bằng phương pháp truyền thống* (tra cứu thủ công) với những vấn đề đơn giản, với kinh nghiệm của các cán bộ tra cứu tốt, nguồn tin không lớn, có phương pháp tra cứu cụ thể, người tìm tin có thể lựa chọn tài liệu / thông tin một cách nhanh chóng, chính xác, tiết kiệm được thời gian, công sức cho cán bộ tra cứu cũng như của người dùng tin.

Thực hiện việc tra cứu truyền thống phải bắt đầu từ các bảng tra cứu - chỉ dẫn (Index), tiến hành từ vấn đề chung đến vấn đề cụ thể. Với các yêu cầu tin đơn giản như hỏi về một tài liệu cụ thể quen thuộc, cán bộ tra cứu có kinh nghiệm nắm vững nguồn tin có trong thư viện, kho tài liệu không lớn, thì yêu cầu tin được đáp ứng một cách nhanh chóng.

Muốn có thông tin chính xác phải có tài liệu hạt nhân.

+ *Khi tra cứu bằng phương pháp tự động hoá* việc tra cứu được tiến hành trên các CSDL, NHDL với các biểu thức tìm tin. Biểu thức tìm tin là sự kết hợp giữa các từ khoá/ từ chuẩn và các toán tử tìm (toán tử Boole/ toán tử logic): OR - Mở rộng phương trình tìm; AND - Thu hẹp phương trình tìm; NOT - Phương pháp loại trừ. Trong quá trình thực hiện việc tra cứu, biểu thức tìm được người tìm tin điều chỉnh cho phù hợp với yêu cầu tin nhằm đạt được kết quả như ý muốn.

1.6. Phân tích kết quả của chiến lược tìm tin

Kết quả của chiến lược tìm tin được thể hiện bằng kết quả dưới nhiều hình thức khác nhau (thông qua việc thực hiện một số chiến lược tìm và biểu thức tìm khác nhau). Trên cơ sở đó kiểm tra, đánh giá các kết quả tìm được và lựa chọn kết quả được coi là phù hợp nhất.

Theo đánh giá chủ quan của cán bộ tra cứu có kinh nghiệm, nếu kết quả nhận được qua tra cứu: quá nhiều hoặc ngược lại, quá ít tài liệu/ thông tin (số liệu, dữ kiện), cần thiết phải phân tích lại yêu cầu tin, thể hiện bằng ngôn ngữ tìm tin, lựa chọn công cụ cho sát thực và xác định lại nguồn tìm.

Kết quả tìm tin được có thể dưới một/ nhiều dạng thức sau:

- Tài liệu gốc/ bản sao: Sách, tạp chí, báo cáo tổng kết, biểu đồ, tài liệu nghe nhìn, đoạn trích, chương, phần của một tài liệu...
- Dữ kiện/ tập hợp dữ kiện.
- Số liệu/ tập hợp các số liệu...

1.7. Biên tập và trình bày thông tin

Biên tập và trình bày thông tin tìm được coi là một bước quan trọng trong chiến lược tra cứu. Nó giúp cho người dùng tin tiếp nhận thông tin một cách thuận tiện, khoa học. Trong các cơ quan thông tin thư viện, thông tin/ tài liệu thường được trình bày dưới nhiều dạng: dạng thông tin thô (Tài liệu gốc) và dạng đã xử lý.

- Danh mục tài liệu,
- Bản thư mục, bài (tóm tắt, tổng quan: thông tin đã được xử lý hình thức và nội dung),
- Tập dữ liệu...

Sắp xếp:

- Theo thứ tự chữ cái tên tác giả/ Nhân vật,
- Theo tiêu đề mô tả,
- Theo môn ngành khoa học,
- Theo đề mục chủ đề,
- Theo thời gian xuất bản tài liệu,
- Theo ngôn ngữ,
- Theo mức độ phức tạp của vấn đề, ...

1.8. Đánh giá tính phù hợp của thông tin nhận được

Khi nhận được tài liệu/ thông tin từ cơ quan thông tin - thư viện, người dùng tin đánh giá tính chính xác, sự phù hợp của thông tin nhận được với các tiêu chí sau:

- Chất lượng,
- Giá trị của thông tin được cung cấp.

Trong trường hợp kết quả tra cứu không phù hợp, quá trình tìm tin phải quay lại từ bước 1: Tìm hiểu yêu cầu tin và phân tích vấn đề đến bước 8: Đánh giá tính phù hợp của thông tin nhận được.

Tuy nhiên đôi khi sự đánh giá của người dùng tin cũng có thể mang tính chủ quan, cần phân tích, đánh giá nguyên nhân không phù hợp của thông tin nhận được. Thông thường do một số nguyên nhân sau:

+ Người dùng tin đưa yêu cầu thông tin không sát đề tài - vấn đề hoặc không biết cách thể hiện yêu cầu tin.

- Người dùng tin không hiểu biết đầy đủ về khả năng của cơ quan thông tin thư viện cũng như nguồn thông tin/ tư liệu mà họ yêu cầu.

- Người dùng tin và cán bộ thông tin chưa hiểu nhau hoặc chưa có quan hệ hợp tác với nhau.

- Người dùng tin đưa ra yêu cầu tin không chính xác, mơ hồ do không xác định được chủ đề: quá rộng hoặc quá hẹp, mức độ sử dụng thông tin nông hoặc sâu quá.

- + Cán bộ xử lý yếu chuyên môn, nghiệp vụ,
- + Bộ máy tra cứu không phong phú, đa dạng.
- + Mạng tin không đáp ứng,

+ Người tìm tin không hiểu đúng câu hỏi do: yếu về trình độ chuyên môn, trình độ ngoại ngữ, trình độ tin học, ...

Quá trình tra cứu tin chỉ có thể kết thúc khi người dùng tin đánh giá: thông tin nhận được phù hợp với yêu cầu tin của họ. Kết thúc quá trình tìm tin, cơ quan thông tin thông tin - thư viện đánh giá kết quả tìm, có dựa trên cơ sở đánh giá của người dùng tin. Việc đánh giá quá trình tìm tin nhằm hai mục đích:

- Làm cho người dùng tin ngày càng thoả mãn yêu cầu.

- Để cải tiến và hoàn thiện hoạt động của cơ quan thông tin thư viện

Để phục vụ tốt cơ quan Thông tin – Thư viện:

- Bổ sung nguồn lực thông tin một cách đều đặn và cập nhật, phù hợp với chức năng, nhiệm vụ của cơ quan thông tin thư viện.

- Tăng cường cơ sở vật chất - kỹ thuật cho cơ quan thông tin - thư viện, đặc biệt những trang thiết bị hiện đại, đáp ứng nhu cầu phát triển khoa học công nghệ trong nước và trên thế giới.

- Nâng cao trình độ của cán bộ thông tin về chuyên môn nghiệp vụ, ngoại ngữ, tin học.

Nếu trong quá trình phục vụ tra cứu, cơ quan thông tin thư viện luôn đáp ứng yêu cầu của người dùng tin, cung cấp thông tin đầy đủ, nhanh chóng, chính xác và cập nhật, thì người dùng tin sẽ tin tưởng và đến với cơ quan nhiều hơn. Vai trò, uy tín cũng như hiệu quả hoạt động của cơ quan được tăng lên.

Tính thích hợp của thông tin nhận được, được đánh giá ở 4 mức độ:

+ Cơ bản

+ Trung bình

+ Đáng chú ý

+ Thờ

Quá trình tra cứu trong thư viện có thể được tiến hành bằng hai cách: *Tra cứu trực tiếp và tra cứu gián tiếp.*

Tra cứu trực tiếp (Direct search): Người dùng tin trực tiếp thực hiện tất cả hoặc hầu hết các thao tác tìm kiếm cùng với sự hỗ trợ đơn giản của chuyên gia thông tin, như: việc định hướng nên tra tìm bằng các công cụ nào để nhận được thông tin cần thiết, cũng như góp ý về cách thức thể hiện yêu cầu tin thông qua chiến lược và lập biểu thức tìm,...

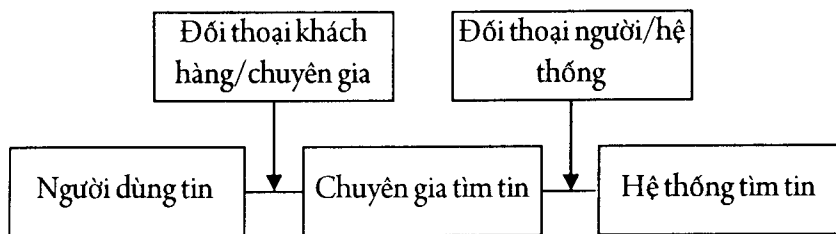
Khi phép tìm không đạt được kết quả như ý muốn, cần phải xác định lại yêu cầu cho rõ hơn, cũng như xem xét lại một số công đoạn như vạch chiến lược tìm, xây dựng biểu thức tìm, xác định nguồn thông tin cấp 2...

Chất lượng của quá trình tra cứu tin phụ thuộc vào sự hiểu biết về nguồn thông tin, bộ máy tra cứu, sự hiểu biết lẫn nhau giữa người dùng tin và chuyên gia thông tin, kỹ năng điều chỉnh chiến lược tìm tin và biểu thức tìm để đạt tới kết quả, nhằm thoả mãn yêu cầu.

Kết thúc quá trình tra cứu, cơ quan thông tin thư viện có thể cung cấp nguồn tin cấp 1 (bản sao tài liệu gốc) hoặc xử lý sâu về nội dung tài liệu: làm tóm tắt, làm tổng quan, lược dịch...

Tra cứu tin gián tiếp (Indirect search): Đa số trường hợp người dùng tin thông qua chuyên gia thông tin để thực hiện việc tra cứu. Chuyên gia tra cứu đóng vai trò là người trung gian giữa người dùng tin và hệ thống tìm tin. Họ có trách nhiệm đảm bảo sự truyền thông có hiệu quả giữa người dùng tin đầu cuối và hệ thống tìm tin. Để đạt được điều đó, chuyên gia tra cứu sẽ phải tìm hiểu một cách cặn kẽ nhu

câu tin, thông qua quá trình phỏng vấn và tham khảo từ người dùng tin. Lược đồ tổng quát của quá trình tra cứu tin gián tiếp có thể được mô tả như trong hình 4.



Hình 4. Sơ đồ quá trình tra cứu tin gián tiếp

Như đã đề cập ở phần trên (các bước của quá trình tra cứu), người tìm tin cần thiết phải đối thoại trực tiếp với người dùng tin để làm rõ một số vấn đề trước khi tiến hành tra cứu như:

- Người dùng tin hiểu như thế nào về yêu cầu tin?
- Xuất phát từ đâu người dùng tin đưa ra yêu cầu?
- Mục đích của yêu cầu phục vụ cho việc gì?
- Loại yêu cầu thông tin nào họ cần?
- Loại tài liệu nào phù hợp với họ?
- Ngôn ngữ có khả năng sử dụng được?
- Hình thức của thông tin được cung cấp?
- Phạm vi chính xác về không gian, thời gian mà kết quả cần bao quát?
- Thời gian mong nhận được kết quả?
- Khi thực hiện quá trình tìm tin, chuyên gia tìm cần có sự hiểu biết và kỹ lưỡng chi tiết về:

- Sự phân bố các nguồn tin.

+ Các cơ quan lưu trữ tài liệu cấp 1: theo loại hình tài liệu, hình thức xuất bản :

Tài liệu công bố có tại:

- Thư viện Quốc gia.
- Hệ thống thư viện công cộng.
- Các thư viện chuyên ngành, đa ngành.
- Thư viện trường học,....

Tài liệu không công bố (chưa được xuất bản: bản thảo khoa học, tư liệu điều tra, tư liệu dịch,...).

- Các trung tâm thông tin - tư liệu.
- Kho lưu trữ nhà nước.
- Trung tâm dịch vụ hỏi và đáp,...
- Các trung tâm phân tích tin (chứa các tài liệu xử lý phân tích tổng thuật, tài liệu tổng quan).
- Cơ quan môi giới và dịch vụ (môi giới thương mại, môi giới thông tin, các cơ quan dịch vụ).
- Hệ thống thông tin hành chính.
- Hệ thống thông tin thống kê kế hoạch,...

Mạng thông tin:

Cung cấp các tài liệu, số liệu nước ngoài, các thông tin đòi hỏi có trình độ cập nhật cao, các thông tin phục vụ tra cứu, trao đổi thông tin nhanh chóng,...

- Mạng LAN,
- Mạng WAN,
- Mạng quốc gia (Intranet),

- Mạng internet,

- *Các công cụ và các nguồn tra cứu thông tin (nguồn tìm).*

- + *Các công cụ và nguồn tra tìm.*

- **Hệ thống mục lục:** Mục lục chữ cái, mục lục phân loại, mục lục chủ đề, mục lục liên hợp, mục lục địa chỉ,...

- **Các loại thư mục (thư mục bậc 2):** Là cơ sở của mọi cuộc tìm kiếm tri thức, giúp người tìm tin có thể khai thác tới nguồn tin cấp 1 được thuận tiện, dễ dàng và đầy đủ.

- **Thư mục trích dẫn (Citation Index):** Là loại bảng tra có tầm quan trọng đặc biệt nhằm phát triển thêm các tư liệu thích hợp với đề tài.

- Thư mục trích dẫn là bảng liệt kê theo vần chữ cái tên tác giả được trích dẫn. Trong từng đề mục (tên tác giả được trích) là danh sách các tác giả trích dẫn cùng một số thông tin bổ sung như các thông tin về nguồn (tên tạp chí, số tạp chí, năm xuất bản của tạp chí, ...) có chứa các bài viết của các tác giả trích dẫn.

- Thư mục trích dẫn nguồn (source Index): Là bảng liệt kê theo vần chữ cái tên các tác giả trích dẫn cùng với các thông tin thư mục đầy đủ của tài liệu/ thông tin trích dẫn.

- Thư mục trích dẫn chủ đề hoán vị (Permuterm Subject Index).

- Là danh sách theo vần chữ cái các thuật ngữ được sử dụng để tạo ra các đề mục chủ đề, và trong mỗi thuật ngữ là theo danh sách theo vần chữ cái các tác giả trích dẫn có bài viết được phản ánh trong chỉ dẫn và đề cập tới nội dung có liên quan tới thuật ngữ này.

- **Danh mục:** Là một bảng liệt kê cho phép xác định được thông tin về một/ một nhóm đối tượng nào đó thuộc các lĩnh vực hoạt động xã hội hoặc khu vực địa lý. Đối tượng được phản ánh trong danh mục có thể:

○ Cá nhân, cơ quan, đơn vị hành chính, đơn vị kinh tế, sản phẩm hàng hoá, nhà sản xuất,...

○ Lĩnh vực hoạt động xã hội: nghiên cứu và triển khai, các ngành dịch vụ, các ngành sản xuất, giáo dục và đào tạo,...

○ Khu vực địa lý: Tỉnh, thành phố, một số tỉnh, thành phố (khu vực, quốc gia, một số quốc gia (khu vực quốc tế), toàn thế giới.

● **Hệ thống phiếu tra cứu dữ kiện:** Là tập hợp các phiếu chứa các thông tin dữ kiện về một vấn đề cụ thể được sắp xếp theo một trật tự xác định.

● **Cơ sở dữ liệu:** Là một tập hợp các bản ghi hoặc có quan hệ logic với nhau và được lưu trữ trên bộ nhớ của máy tính.

Trong khi tìm hiểu câu hỏi phải thận trọng, tránh câu hỏi bị nhiều hoặc im lặng. Trong thực tế, có nhiều nguyên nhân khiến các chuyên gia thông tin không nắm bắt được đúng yêu cầu thực sự. Nguyên nhân có thể là do:

Người dùng tin không chi rõ và không muốn lộ ra mục đích thực sự trong câu hỏi của họ.

Ví dụ:

Có trường hợp người dùng tin đến thư viện hỏi những thông tin liên quan tới cây thuốc phiện (giống, đất trồng, chăm sóc, thu hoạch,...).

Mục đích: Muốn hỏi trồng cây thuốc như thế nào? nhưng việc trồng cây thuốc phiện bị cấm, do đó họ không thể nói rõ mục đích với những câu hỏi trực tiếp tới vấn đề, mà hỏi từng ý liên quan vấn đề vào các thời gian khác nhau. Trong trường hợp này cán bộ thông tin phải khéo léo đưa ra những câu hỏi có, để NĐT thể hiện mục đích đích thực của mình và cơ quan thông tin cung cấp thông tin phù hợp.

2. QUÁ TRÌNH TÌM TIN TỰ ĐỘNG HÓA

2.1. Các bước của quá trình tìm tin tự động hoá

Quá trình tìm tin tự động hoá, dù đó là quá trình có hoặc không có chuyên gia tìm tin trung gian cũng có thể được chia thành nhiều bước theo một trình tự hợp lý. Về cơ bản quá trình tìm tin tự động hoá gần giống như các bước tra cứu tin cơ bản. Tuy nhiên, do đặc thù cần phải sử dụng những hệ thống tìm tin tự động hoá, các bước của quá trình tìm tin tự động hoá có thể có một số khác biệt so với tra cứu thông tin theo phương pháp truyền thống hoặc thủ công.

Tìm tin tự động hoá là một quá trình truyền thông giữa người có nhu cầu tìm tin và hệ thống tìm tin. Quá trình tìm tin tự động hoá ở những loại hình tìm tin khác nhau có thể khác nhau. Quá trình tìm tin gián tiếp thông qua chuyên gia tìm tin cũng không giống với quá trình tìm tin của người dùng tin trực tiếp. Quá trình tìm tin trên những hệ thống cục bộ, CD-ROM có thể khác với quá trình tìm tin trực tuyến.

Về cơ bản quá trình tìm tin tự động hoá có thể bao gồm các bước sau:

- (1) Phân tích nhu cầu tin.
- (2) Lựa chọn nơi tìm tin.
- (3) Xác định các khái niệm và thuật ngữ tìm.
- (4) Xác định chiến lược tìm.
- (5) Xây dựng biểu thức tìm.
- (6) Truy cập hệ thống và thực hiện lệnh tìm.
- (7) Đánh giá sơ bộ kết quả tìm.
- (8) Tiếp tục thực hiện lệnh hoặc hiệu chỉnh lệnh tìm.

Một số bước trong quá trình tìm tin tự động hoá được giới thiệu chi tiết hơn trong Chương IV (Tìm tin tự động hoá).

2.2. Các bước của tìm tin trực tuyến

Tìm tin trực tuyến là một dạng tìm tin tự động hoá. Do đặc thù là tìm tin trực tuyến thường liên quan đến các dịch vụ thông tin trực tuyến có tính thương mại, mọi thao tác với cơ sở dữ liệu đều bị tính chi phí (chi phí cho số từ khoá, chi phí cho câu hỏi, chi phí cho xem biểu ghi...) nên các bước tìm tin trực tuyến đòi hỏi được chuẩn bị hết sức kỹ càng.

Tác giả Stephen Harter cho rằng tra cứu trực tuyến bao gồm 12 bước như sau:

- (1) Nghiên cứu nhu cầu tin.
- (2) Xác định mục tiêu tìm kiếm.
- (3) Lựa chọn cơ sở dữ liệu/ngân hàng dữ liệu.
- (4) Xác định khái niệm của yêu cầu tìm và quan hệ giữa chúng.
- (5) Lựa chọn chiến lược tổng quát.
- (6) Lựa chọn thuật ngữ và phương án thay thế/dự phòng.
- (7) Xác định các trường giới hạn trong tra tìm.
- (8) Xây dựng lệnh tìm.
- (9) Chuẩn bị phương án dự phòng.
- (10) Truy cập hệ thống và thực hiện theo lệnh tìm được lập.
- (11) Đánh giá sơ bộ kết quả tìm kiếm trung gian.
- (12) Tiếp tục thực hiện lệnh hoặc hiệu chỉnh.

Việc chuẩn bị kỹ càng từng bước sẽ đảm bảo hiệu quả tra cứu tin cao, giảm chi phí truy cập và làm việc với cơ sở dữ liệu.

3. ĐÁNH GIÁ HIỆU QUẢ TRA CỨU

3.1. Tiêu chuẩn đánh giá

Để đánh giá kết quả hoạt động tra cứu phải dựa vào:

- **Tính kịp thời:**

Tính kịp thời của thông tin được cung cấp phụ thuộc vào thời gian trả lời yêu cầu tin. Trong nền kinh tế thị trường, tính kịp thời của thông tin đóng vai trò hết sức quan trọng, nhiều khi đóng vai trò quyết định vì tính cạnh tranh trong hoạt động nghiên cứu, sản xuất, thương mại và trong rất nhiều trường hợp: ai có thông tin người đó sẽ chiến thắng.

- **Giá thành (chi phí):**

Ở Việt Nam vấn đề giá thành nói riêng, cũng như kinh tế nói chung trong hoạt động của các cơ quan thông tin - thư viện vẫn đang là vấn đề mới, chưa có sự hướng dẫn, chỉ đạo thống nhất, mỗi cơ quan tùy theo điều kiện cụ thể, đặc thù riêng của mình để tính chi phí dịch vụ tra cứu, tuy nhiên các cơ quan thông tin thư viện cũng đã dựa vào một số tiêu chí sau để tính:

+ Công sức của cán bộ bỏ ra trong quá trình tra cứu thông tin.

+ Sử dụng hệ thống tìm tin của cơ quan thông tin thư viện hoặc kết nối với mạng thông tin trong nước và quốc tế.

+ Hình thức của sản phẩm đầu ra.

- + Khả năng cung cấp tài liệu gốc/ tài liệu sao.
- **Chất lượng cung cấp thông tin cho người dùng tin gồm:**
 - + Độ tin cậy,
 - + Mức độ bao quát chủ đề,
 - + Tính chính xác,
 - + Mức độ đầy đủ: số lượng tài liệu / thông tin tìm được.

Việc đánh giá theo ba tiêu chuẩn trên được tiến hành một cách song song, nhưng thông thường người ta quan tâm nhất tới chất lượng thông tin được cung cấp cho người dùng tin. Vì đây là yếu tố chính, quan trọng nhất để đánh giá tính phù hợp của thông tin nhận được.

Tuy nhiên, ngoài các tiêu chuẩn trên còn có một số cách đánh giá khác có liên quan tới hiệu quả của quá trình tra cứu.

- Đánh giá các công cụ tra cứu:

- + Mức độ bao quát thông tin, bao phủ đề tài của công cụ tra cứu,
- + Mức độ chính xác,
- + Tiện sử dụng,
- + Tính đa dạng, phong phú ...

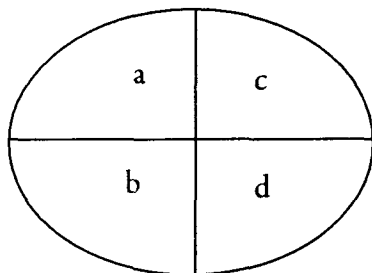
Ví dụ:

Hệ thống mục lục.

- ✓ Khả năng tra cứu được thông tin/ tài liệu thích hợp trong thời gian ngắn nhất.
- ✓ Cung cấp nhiều điểm tiếp cận cho mỗi tài liệu.
- ✓ Chi tiết hoá các hệ thống tiêu đề, đề mục, hệ thống chi chõ,... trong hệ thống mục lục.

3.2. Các chỉ số đánh giá hiệu quả tra cứu thông tin

Hiệu quả tra cứu được xác định bằng một số công thức tính. Các công thức đó được thể hiện qua các chỉ số: Mức độ đầy đủ, mức độ chính xác, tính kịp thời, giá thành,...



Để tính được các chỉ số cần xác định được: số tài liệu thích hợp, không thích hợp, tài liệu tìm được và không tìm được ở trong kho tư liệu.

Giả sử ta có tập hợp $a+b+c+d$ là tổng số tài liệu có trong kho hoặc tổng số biểu ghi trong cơ sở dữ liệu.

	Tài liệu thích hợp	T/liệu không thích hợp	
Tài liệu tìm được	A	B	A+B
T/Liệu không tìm được.	C	D	C+D
Tổng số tài liệu	A+C	B+D	A+B+C+D

Trong đó:

- Tập hợp **a**: Tài liệu thích hợp tìm được.
- Tập hợp **b**: Tài liệu không thích hợp tìm được (tạp tin, nhiễu tin).
- Tập hợp **c**: Tài liệu thích hợp không tìm được (mất tin).
- Tập hợp **d**: Tài liệu không thích hợp, không tìm được.

Tập hợp $a+b+c+d$ thể hiện bằng ma trận như sau:

Dựa vào ma trận trên ta có thể tính được:

Mức độ đầy đủ:

Kí hiệu là R : biểu thị bằng tỉ số giữa tài liệu thích hợp tìm được trên tổng số tài liệu thích hợp có trong cơ quan thông tin thư viện:

$$R = \frac{a}{a + c}$$

Mức độ chính xác:

Kí hiệu là P : biểu thị bằng tỉ số giữa tài liệu thích hợp tìm được trên tổng số tài liệu tìm được.

$$P = \frac{a}{a + b}$$

Tính kịp thời:

Kí hiệu là O - biểu thị bằng tổng số chi phí thời gian tra tìm cho số lần tìm.

$$O = \sum_{i=1}^{i=n} t_i p_i \quad i = 1, 2, \dots, n - \text{số lượng lần tra tìm}$$

Giá thành:

Kí hiệu là C - biểu thị chi phí tiền và vật chất cho tra tìm một tài liệu. Nó được tính bằng chi phí bỏ ra để tra tìm toàn bộ tài liệu tìm được trong khoảng thời gian nhất định.

$$C = \sum_{i=1}^{i=n} \frac{3nmi}{D_B} \quad i = 1, 2, \dots, n - \text{số lượng lần tra tìm}$$

Hệ số lao động:

Kí hiệu là C - Biểu thị bằng tổng số chi phí lao động bỏ ra cho việc tra tìm một tài liệu.

$$T = \sum_{i=1}^{i=n} T_p \quad i = 1, 2, \dots, n - \text{số lượng lần tra tìm}$$

Hệ số lựa chọn:

Kí hiệu là J - Thể hiện số lượng công việc đã thực hiện trong quá trình tra tìm. Hệ số lựa chọn được tính bằng tổng số tài liệu đã tìm được trên tổng số tài liệu có trong kho

$$J = \frac{a + b}{a + b + c + d}$$

Hệ số kịp thời, giá thành và hệ số lao động của quá trình tra cứu thông tin phụ thuộc vào:

- Tổ chức quá trình tra tìm.
- Sự hoàn thiện cơ sở vật chất, kỹ thuật của cơ quan thông tin - thư viện.
- Trình độ chuyên môn, nghiệp vụ của cán bộ tra cứu.
- Quan hệ giữa cơ quan thông tin thư viện và người dùng tin.

Bằng các tiêu chuẩn và chỉ số trên, cơ quan thông tin thư viện có thể phân tích, đánh giá được hoạt động tra cứu nói riêng và hoạt động thông tin nói chung để điều chỉnh, cải tiến công tác của cơ quan mình, đồng thời dự báo được sự phát triển của của công tác này, nhằm đáp ứng nhu cầu đa dạng, phong phú của người dùng tin.

Chương 3

Phương pháp tra cứu thông tin truyền thống

Tra cứu thông tin truyền thống (Tra cứu thủ công) là phương pháp mà quá trình tra cứu được tiến hành thông qua bộ máy tra cứu/ các phương tiện và công cụ truyền thống (thủ công). Trong các cơ quan thông tin thư viện, người dùng tin có thể: Hỏi về một tài liệu cụ thể; Hỏi một phần, một chương trong tuyển tập/ cuốn sách bất kỳ; Hỏi một tác giả cụ thể có những tác phẩm nào trong thư viện; Hỏi về một vấn đề, đề tài; Hỏi các số liệu thống kê, thông số kỹ thuật liên quan đến loại máy móc thiết bị nào đó,... Các dạng câu hỏi trên ta có thể chia ra thành các dạng tra cứu sau: tra cứu thông tin thư mục (tra cứu tư liệu) và tra cứu thông tin dữ kiện (tra cứu dữ kiện).

1. TRA CỨU THÔNG TIN THƯ MỤC

Tra cứu thông tin thư mục (Tìm tin tư liệu) là quá trình xác định và tách ra khỏi nguồn tìm kiếm các tài liệu tương ứng với yêu cầu tin theo các dấu hiệu cho trước như: tên tác giả, tên tài liệu, tên người dịch, tên nhân vật, địa danh, môn ngành tri thức, chủ đề, từ khoá... Kết

quả của quá trình tìm tin là tài liệu gốc (bản sao tài liệu gốc), hoặc thông tin về tài liệu.

Ví dụ:

Tìm tài liệu “Thư mục học đại cương”,

“Tin học trong hoạt động thông tin - thư viện”.

Hoặc những *thông tin về tài liệu như*: Tác giả, dịch giả, người chủ biên, người hiệu đính, lần xuất bản, nơi xuất bản, nhà xuất bản, năm xuất bản, thời gian nộp lưu chiểu,...

Ví dụ:

Tìm tên tác giả của tài liệu: “Định chủ đề tài liệu”?

Năm xuất bản của tài liệu: “Tin học trong hoạt động thông tin - thư viện”?...

Để tìm tài liệu cần phải dựa vào bộ máy tra cứu.

1.1. Nguồn tra cứu chính

- Hệ thống mục lục:

+ *Mục lục chữ cái.*

+ *Mục lục phân loại.*

+ *Mục lục chủ đề.*

+ *Mục lục địa chỉ.*

- Các bộ phiếu tra cứu

+ *Bộ phiếu tra cứu chính.*

+ *Bộ phiếu tra cứu chuyên đề.*

- **Ấn phẩm thông tin - thư mục:**

- + Ấn phẩm thông tin thư mục.
- + Ấn phẩm thông tin tóm tắt.
- + Ấn phẩm thông tin hỗn hợp.
- + Thư mục chuyên đề.
- + Thư mục giới thiệu.
- + Thư mục địa chí...

1.2. Phương pháp tra cứu

1.2.1. Tra cứu qua hệ thống mục lục thư viện

Dạng tra cứu thông tin này phải dựa vào các thuật ngữ chính xác: đó là các từ mô tả như: Tên tác giả, tên tài liệu, nhà xuất bản, địa danh.. hoặc các dấu hiệu đặc trưng cho yêu cầu như: Đề mục chủ đề, môn ngành tri thức/ lĩnh vực khoa học và thường bắt đầu từ những khái niệm quan trọng nhất, sau đó có thể thu hẹp (mở rộng) phương pháp tìm với sự giúp đỡ của các hệ thống tra cứu chi chỗ qua lại, các quan hệ giữa các thuật ngữ với nhau.

Hệ thống mục lục phiếu (Hệ thống mục lục thư viện) là nguồn tra cứu quan trọng và thông dụng trong các thư viện, vì nó liệt kê một cách hệ thống tất cả tài liệu có ở trong thư viện, cho phép tìm được vị trí (địa chỉ) lưu trữ tài liệu trong kho của một/một số cơ quan thông tin thư viện, với điều kiện người tìm tin biết một/ các thông tin về tài liệu như: tác giả của tài liệu, tên tài liệu, nội dung của tài liệu: chủ đề, hay môn ngành khoa học. Tùy thuộc vào cách tổ chức mục lục mà người ta sử dụng tới các nhóm thông tin về tài liệu khác nhau để tra cứu trong mục lục tương ứng.

- **Mục lục chữ cái (MLCC)** cho phép tìm tài liệu theo tên tác giả, tên tài liệu, tổng số tài liệu của một tác giả được xuất bản và lưu trữ ở kho của thư viện, cơ quan thông tin.

Ví dụ:

✓ Ai là tác giả của tiểu thuyết: "Chú nhóc đen"?

Phương pháp tra cứu:

MLCC: Văn C - CH: Chú nhóc đen

Tác giả: R. Raito.

✓ Nguyễn Thụy Ứng dịch tiểu thuyết nào xuất bản năm 1983?

Phương pháp tra cứu:

MLCC: Văn: N - Nguyễn Thụy Ứng: Người dịch.

Tiểu thuyết: Sông Đông êm đêm/ M. Sôlôkhốp, Nguyễn Thụy Ứng: dịch.- Tiểu thuyết. (Bản dịch mới)

Tiểu thuyết: Sông Đông êm đêm

✓ Cho biết nhà xuất bản tài liệu "Món ăn Huế"?

Phương pháp tra cứu:

MLCC: Văn: M - Món ăn Huế.

Món ăn Huế do nhà xuất bản Phụ nữ XB.

Nhà xuất bản: Phụ nữ

- **Mục lục phân loại (MLPL):** cho phép tìm tài liệu theo môn ngành khoa học (lĩnh vực hoạt động). Mỗi mục lục phân loại có cách tổ chức tổ chức riêng tùy theo Bảng phân loại được áp dụng ở cơ quan thông tin thư viện cụ thể.

Thông thường mỗi cơ quan thông tin thư viện chi tổ chức phân loại và xây dựng mục lục phân loại theo một hoặc hai bảng, song trên

thực tế có không ít thư viện tồn tại đồng thời nhiều mục lục phân loại (đa phần do hoàn cảnh lịch sử để lại). Ví dụ ở *Thư viện Quốc gia Việt Nam* tồn tại:

Mục lục phân loại:

- + Theo Bảng Trung tiêu hình của Trung Quốc.
- + Theo Bảng phân loại Dừng cho các thư viện khoa học tổng hợp.
- + Theo bảng phân loại BBK rút gọn.
- + Theo bảng phân loại DDC rút gọn.

Thư viện Khoa học và Kỹ thuật Trung ương (thuộc Trung tâm Thông tin KH & CN Quốc gia) có:

Mục lục phân loại.

- + Theo bảng phân loại Trung tiêu hình của Trung Quốc.
- + Theo bảng phân loại BBK rút gọn,

Đối với người tra cứu, để tra tìm được tài liệu theo môn ngành tri thức phải nắm chắc cơ cấu từng bảng phân loại, đặc điểm của chúng mới có khả năng tra tìm một cách nhanh chóng và đầy đủ, đáp ứng yêu cầu của người dùng tin.

Ví dụ:

- ✓ Tìm các tài liệu về khai thác Amiăng?

Thư viện sử dụng Bảng Phân loại dùng cho các thư viện khoa học tổng hợp (năm 2002).

Phương pháp tra cứu:

MLPL, Mục: 6C1: Ngành khai mỏ,

.....

6C1.8: Khai thác các khoáng sản phi kim loại.

6C1.82: Khai thác mỏ vật liệu xây dựng và vật liệu làm đường giao thông.

Khai thác Amiăng thuộc: Khai thác mỏ vật liệu xây dựng và vật liệu làm đường giao thông.

Khai thác Amiăng: 6C1.82

Bảng Phân loại Thư viện - Thư mục (BBK) với câu hỏi trên sẽ tìm:

G: ngành mỏ.

G 31/34: Khai thác mỏ từng loại.

G 34: Khai thác mỏ khoáng sản rắn phi quặng (phi kim loại).

G 342: Khai thác mỏ vật liệu xây dựng và làm đường.

Khai thác Amiăng: G342

✓ Tìm các tài liệu về *Bệnh run tay*?

Phương pháp tra cứu:

Tương tự như cách tra cứu trên: sử dụng Bảng phân loại dùng cho các thư viện khoa học tổng hợp của Thư viện Quốc gia (năm 2002).

MLPL, 61: Y học. Y tế

617.9 Bệnh lý thần kinh

617.919: Các bệnh khác của hệ thần kinh

Bệnh run tay: 617.919

BBK:

N: Y tế. Các khoa học y học

N 61 / 64 Bệnh lý thần

N 62: Bệnh lý thần kinh

N 629: Các bệnh khác của hệ thần kinh

Bệnh run tay: N629

- **Mục lục chủ đề (MLCD)**: cho phép tìm được tin theo các đề mục chủ đề.

Tại các thư viện, MLCD được xây dựng trên cơ sở **Bảng đề mục chủ đề của Thư viện Quốc hội Mỹ (LCSH)**, **Bảng đề mục chủ đề của Thư viện Quốc gia Pháp (RAMEAU)** hay Bảng đề mục chủ đề do thư viện tự xây dựng như Bảng phân loại chủ đề của Đại học Y Hà Nội.

Ví dụ:

✓ Tìm các giáo trình về tiêu hoá có tại thư viện Trường Đại học Y Hà Nội.

Phương pháp tra cứu:

Đề mục chủ đề,

Chủ đề: Tiêu hoá: T 8,

Giáo trình: 01

Giáo trình tiêu hoá: T 8. 01

✓ Tìm các tài liệu để cập tới Phật giáo Việt Nam.

Tại Thư viện Quốc gia Việt Nam có MLCD (sách từ 1954 - 1960) với hơn 2000 đề mục.

Phương pháp tra cứu:

Đề mục chủ đề,

Chủ đề: Phật giáo: Việt Nam.

Các tài liệu để cập tới *Phật giáo Việt Nam* có ở **Chủ đề**: Phật giáo: Việt Nam. Tuy nhiên sách chỉ có đến 1960, muốn tra tìm tài liệu mới hơn có thể tra tìm trong MLPL (BBK):

X tôn giáo. Chủ nghĩa vô thần

X 3 Từng loại tôn giáo

X 3.... Phật giáo

Phật giáo Việt Nam X 3....(2)

Qua hệ thống mục lục có thể lựa chọn tài liệu bằng việc kết hợp và sử dụng các thông tin khác nhau về tài liệu mà chúng không được sử dụng làm dấu hiệu để xây dựng mục lục như:

+ Các thông tin về xuất bản: Nơi xuất bản, nhà xuất bản, năm xuất bản.

+ Các thông tin về đặc điểm số lượng của tài liệu: số trang, số khổ, số liệu minh hoạ,...

1.2.2. Tra cứu qua ấn phẩm thông tin, các bản thư mục và mục lục in

Dạng tra cứu này được tiến hành giống như tra cứu qua hệ thống mục lục. Tốc độ tra cứu ở đây phụ thuộc vào mức độ phức tạp của biểu thức phản ánh yêu cầu tin, cách tổ chức các hình thức lưu trữ và khả năng tra cứu thông tin của các công cụ được sử dụng.

Mọi việc tra cứu đều được bắt đầu từ những thông tin quan trọng nhất liên quan tới nội dung hoặc hình thức cụ thể của tài liệu, sau đó những thông tin khác cũng cần được kiểm tra để yêu cầu được thoả mãn một cách chính xác.

Trong ấn phẩm thông tin hay bản thư mục, thời gian và khả năng tra tìm thông tin phụ thuộc vào nguyên tắc sắp xếp các biểu ghi thư mục, các bài tóm tắt và hệ thống các bảng tra cứu, bố trợ của thư mục/ ấn phẩm, nếu hệ thống các bảng tra cứu, bố trợ phong phú, đa dạng

thì khả năng truy nhập được thông tin theo nhiều khía cạnh khác nhau một cách nhanh chóng, đầy đủ.

Trong khi tiến hành tra cứu, nhiệm vụ của người tìm tin là xác định dạng tìm tin đó thuộc loại nào?

- **Theo hình thức:** Tìm theo tên tác giả, tên tài liệu, các yếu tố xuất bản;

- **Theo nội dung:** Tìm theo chủ đề, môn ngành khoa học, từ khóa/ các dạng thông tin khác thích hợp với thuật ngữ tìm của yêu cầu, trên cơ sở đó xác định các bảng tra cứu, chỉ dẫn phù hợp với yêu cầu.

Để thực hiện việc tìm tin có hiệu quả, với những vấn đề mới chưa cập nhật và lĩnh vực chuyên môn không thật am hiểu, trước hết phải tham khảo các tài liệu để nắm khái niệm liên quan tới các thuật ngữ, vấn đề hoặc lĩnh vực cần tìm trong tài liệu tra cứu như: bách khoa thư, từ điển bách khoa, từ điển và sổ tay tra cứu chuyên ngành hoặc các sách giáo khoa, giáo trình, tổng quan,...

Trong thư mục, mục lục in (dưới dạng sách) hoặc ấn phẩm thông tin thông thường có các bảng tra cứu / chỉ dẫn như:

- + Bảng tra tác giả (Author Index).
- + Bảng tra chủ đề (Subject Index).
- + Bảng tra từ khoá (Keyword Index).
- + Bảng tra văn bằng phát minh (Patent Index).
- + Bảng tra công thức (Formule Index).
- . Bảng tra công thức chung.¹
- . Bảng tra công thức hợp chất hệ vòng.
- + Bảng tra khái niệm (Concept Index).

+ Bảng tra chung (Semi-annual Commulative Index (nửa năm), Commulative Index (cả năm),...

Khi tìm tin phải xác định rõ một hoặc một vài yếu tố như: lĩnh vực khoa học, chủ đề, từ khoá, tên tác giả, tên tài liệu / văn bằng phát minh, công thức hoá học.... liên quan tới yêu cầu, sau đó xác định các bảng tra tương ứng để tìm. Mỗi loại ấn phẩm thông tin có các bảng tra cứu phù hợp với nội dung tài liệu mà chúng phản ánh.

Tham khảo một số ấn phẩm thông tin nước ngoài và Việt Nam sau:

1. Tạp chí tóm tắt hoá học Mỹ (Chemical Abstracts)

Cơ quan xuất bản	Hội hoá học Mỹ (Ohio). 1907-
Định kỳ xuất bản	Hàng tuần + các bảng chỉ dẫn. Dưới dạng giấy + CD-ROM.
Phạm vi bao quát tài liệu	- Giới thiệu các tài liệu khoa học kỹ thuật thế giới, các bài báo, tạp chí, công trình hội nghị, luận án, báo cáo, văn bằng phát minh về hoá học và công nghệ hoá học.
Cách trình bày bài tóm tắt	Mã số (số tập, số thứ tự bài). Tên tư liệu bằng tiếng Anh. Tác giả (nơi công tác của tác giả). Tên nguồn dẫn tư liệu (ngôn ngữ của tư liệu nguyên bản). Tóm tắt.
Phân loại	Theo phân loại khoa học.
Bảng chỉ dẫn/ Bảng tra cứu	- Trong mỗi số có các bảng chỉ dẫn. + Từ khoá. + Văn bằng phát minh.

	<ul style="list-style-type: none"> + Tác giả. - <i>Bảng chỉ dẫn chung</i> (6 tháng) (Semi - annual commulative). + Chủ đề: chung, các chất. + Công thức: công thức chung, công thức hợp chất hệ vòng. + Chi dẫn tác giả. + Chi dẫn văn bằng phát minh.
Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: + Theo từng vấn đề trong mục lục của ấn phẩm. + Theo bảng chỉ dẫn kèm theo mỗi số (dựa vào dấu hiệu tìm kiếm). - Tra cứu tin hồi cố + Theo các bảng chỉ dẫn chung (dựa vào các dấu hiệu tìm kiếm). + Theo từng số.

2. Tạp chí tóm tắt sinh vật Mỹ (Biological ABSTRACTS)

Cơ quan xuất bản	Trung tâm dịch vụ thông tin sinh vật (Philadelfia). 1927-
Định kỳ xuất bản	24 số / năm + một bản chỉ dẫn.
Phạm vi bao quát tài liệu	<ul style="list-style-type: none"> - Giới thiệu về tài liệu khoa kỹ thuật thế giới về sinh vật học và các ngành tri thức có liên quan. - Hàng năm giới thiệu khoảng 70.000 bài tóm tắt.

<p>Cách trình bày bài tóm tắt</p>	<ul style="list-style-type: none"> - Số thứ tự. Tên tác giả (địa chỉ cơ quan của tác giả). - Tên tài liệu bằng tiếng Anh. Nguồn dẫn tài liệu (ngôn ngữ của tài liệu nguyên bản). Tóm tắt.
<p>Phân loại</p>	<ul style="list-style-type: none"> - Theo phân loại khoa học.
<p>Bảng chỉ dẫn/ bảng tra</p>	<ul style="list-style-type: none"> - <i>Trong mỗi số</i> + Bảng chỉ dẫn tác giả (Author Index). + Bảng tra phân loại sinh vật (Biosystemetic Index). + Bảng tra khái niệm (Concept Index). + Bảng tra chủ đề (Subject Index). - <i>Bảng chỉ dẫn chung</i> (Semi - annual commulative): 6 tháng.
<p>Phương pháp tra cứu tin</p>	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: + Dùng các bảng tra trong mỗi số. + Theo mục lục của ấn phẩm. - Tra cứu tin hồi cố: + Dùng các bảng chỉ dẫn chung. + Tìm theo từng số riêng lẻ.

3. Tạp chí tóm tắt nông nghiệp AGRINDEX

Cơ quan xuất bản	- Hệ thống Thông tin Quốc tế về nông lương, Liên hợp quốc (Viên)
Định kỳ xuất bản	<ul style="list-style-type: none"> - 12 số/năm + 1 bản chỉ dẫn chung. - Dưới dạng giấy, Microfilm và CD-ROM.
Phạm vi bao quát tài liệu	<ul style="list-style-type: none"> - Giới thiệu về tư liệu khoa kỹ thuật thế giới liên quan đến nông nghiệp. - Hàng năm giới thiệu khoảng 100.000 bài tóm tắt.
Cách trình bày bài tóm tắt	<ul style="list-style-type: none"> - Mã số (năm, số thứ tự). Tên tư liệu nguyên bản [Tên tư liệu dịch sang tiếng Anh]. [Mã ngôn ngữ]. Tác giả (địa chỉ nơi công tác của tác giả). Nguồn dẫn tư liệu. Từ khoá theo AGROVOC.
Phân loại	<ul style="list-style-type: none"> - Theo phân loại khoa học.
Bảng chỉ dẫn/ tra cứu	<ul style="list-style-type: none"> - Trong mỗi số: tra cứu theo: <ul style="list-style-type: none"> + Bảng tra tác giả (cá nhân). + Bảng tra tác giả (tập thể). + Bảng tra các báo cáo. + Bảng tra chủ đề. - Bảng chỉ dẫn chung (Agrindex Commulative Index) - 1 năm: tra cứu theo: <ul style="list-style-type: none"> + Tác giả cá nhân. + Tác giả tập thể. + Chủ đề.
Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: <ul style="list-style-type: none"> + Theo mục lục của ấn phẩm.

	<ul style="list-style-type: none"> + Dùng các bảng tra kèm theo mỗi số - Tra cứu tin hồi cố + Dùng bảng tra chung + Theo mỗi số lẻ
--	--

4. Tạp chí tóm tắt của Nga (Реферативный журнал)

Cơ quan xuất bản	Viện thông tin khoa học kỹ thuật toàn liên bang VINITI .1953-
Định kỳ xuất bản	12 số, 24 số/ năm, có 1 bảng tra cứu chủ đề và 1 bảng tra cứu tác giả chung. Xuất bản dưới dạng giấy, Microfilm và CD-ROM.
Phạm vi bao quát tài liệu	<ul style="list-style-type: none"> - Giới thiệu tài liệu khoa học tự nhiên, khoa học kỹ thuật và khoa học ứng dụng. - Phản ánh hơn 17.000 tạp chí, 19.000 sách, hơn 230.000 văn bằng phát minh của 130 nước, bằng 66 thứ tiếng khác nhau trên thế giới. - Hàng năm giới thiệu khoảng 1.300.000 bài tóm tắt dưới dạng: Các tập liên hợp, tập riêng biệt và các tập lẻ về các bộ môn khoa học. Mỗi loại đều có mã số riêng.
Cách trình bày bài tóm tắt	<ul style="list-style-type: none"> - Mã số (số thứ tự tập, mã loại, số thứ tự bài). Tên tài liệu dịch sang tiếng Nga (tên nguyên bản)/ tên tác giả// tên nguồn dẫn tư liệu... mã ngôn ngữ. Tóm tắt.

Phân loại	<ul style="list-style-type: none"> - Khung đề mục của hệ thống khoa học kỹ thuật Quốc gia. Danh mục đề mục in trong số một hàng năm.
Bảng tra cứu/ chỉ dẫn	<ul style="list-style-type: none"> - Trong mỗi số: <ul style="list-style-type: none"> + Bảng tra văn bản phát minh xếp theo số thứ tự. + Bảng tra tác giả. + Bảng tra chủ đề. - Hàng năm: <ul style="list-style-type: none"> + Bảng tra chủ đề. + Bảng tra công thức hoá học. + Bảng danh mục nguồn trích dẫn tài liệu + Bảng tra tác giả. + Bảng phân loại sinh vật.
Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: <ul style="list-style-type: none"> + Theo đề mục ở mỗi số. + Theo các bảng tra trong mỗi số. - Tra cứu tin hồi cố: <ul style="list-style-type: none"> + Dùng bảng tra cứu chung của năm. + Dùng bảng tra trong từng số lẻ.

5. Tạp chí Thông tin tín hiệu của Pháp (Bulletin signalitique)

Cơ quan xuất bản	Trung tâm thông tin KHKT Quốc gia Pháp (CNRS)
Định kỳ xuất bản	12 số/ năm + 1 bản chỉ dẫn chung dưới dạng giấy + CD-ROM.
Phạm vi bao quát tài liệu	<p>Giới thiệu các tài liệu KHKT thế giới, các bài báo của 9.000 loại ấn phẩm định kỳ; Công trình hội nghị, luận án, báo cáo,...</p> <p>Hàng năm giới thiệu khoảng 800.000 bài mô tả thư mục.</p> <p>- Có 63 loại. Mỗi loại có mã số riêng giới thiệu các tài liệu cụ thể về một ngành cụ thể trong lĩnh vực khoa học chính xác, công nghệ, các khoa học trái đất, sinh y học, các ngành kỹ thuật.</p>
Cách trình bày	- Mã số (năm, mã loại, số thứ tự bài). Tên tài liệu (nguyên bản). Tên tác giả (địa chỉ nơi công tác của tác giả). Nguồn dẫn (nguồn trích dẫn) tài liệu. Từ khoá hoặc thông tin chú thích.
Phân loại	- Theo phân loại khoa học.
Bảng tra cứu/ chỉ dẫn	<p>- Trong mỗi số:</p> <p>+ Bảng tra tác giả. + Bảng tra chủ đề.</p> <p>+ Theo ký hiệu phân loại.</p> <p>- Hàng năm:</p> <p>+ Bảng tra chủ đề. + Bảng tra luận án.</p> <p>+ Bảng tra báo cáo. + Bảng tra biên bản hội nghị</p> <p>+ Danh mục nguồn dẫn tài liệu.</p>

Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: + Tra theo bảng chỉ dẫn cho mỗi số. + Theo từng vấn đề của khung phân loại. - Tra cứu tin hồi cố: Tra theo các bảng tra chung.
-------------------------	--

6. Tạp chí tóm tắt tài liệu khoa học và công nghệ Việt Nam

Cơ quan xuất bản	Trung tâm Thông tin Khoa học và Công nghệ Quốc gia (1977-)
Định kỳ xuất bản	<ul style="list-style-type: none"> - 12 số / năm & các bảng tra (trong mỗi số). - Bảng tra chung: tác giả (cả năm).
Phạm vi bao quát tài liệu	- Giới thiệu các dạng tài liệu khoa học công nghệ Việt Nam: tạp chí, sách, tài liệu hội nghị, luận án tiến sĩ và tiến sĩ khoa học bảo vệ trong nước.
Cách trình bày bài tóm tắt	<ul style="list-style-type: none"> - Chi số khung đề mục quốc gia. - Mã số tài liệu, số thứ tự. Tác giả. Tên tài liệu/ báo cáo/ luận án; tóm tắt. từ khoá. Nơi lưu trữ tài liệu; nguồn tài liệu, số liệu xuất bản, chỉ số ISSN của tạp chí.
Phân loại	- Theo khung đề mục quốc gia
Bảng tra cứu/ chỉ dẫn	<ul style="list-style-type: none"> - Trong mỗi số: + Bảng tra tác giả. + Bảng tra từ khoá. + Bảng tra địa lý. - Hàng năm: Bảng tra chung : Bảng tra tác giả.

Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: + Theo từng vấn đề của khung đề mục. + Dùng bảng tra cứu - chỉ dẫn trong mỗi số. - Tra cứu tin hồi cố: + Bảng tra chung: Tên tác giả. + Theo bảng tra từng số lẻ của các năm.
-------------------------	--

7. Thư mục Quốc gia Việt Nam

Cơ quan xuất bản	Thư viện Quốc gia Việt Nam (Hà Nội) 1957-
Định kỳ xuất bản	<ul style="list-style-type: none"> - Hàng tháng: 12 số/ năm. - Hàng năm: 1 số/ năm.
Phạm vi bao quát tài liệu	- Giới thiệu các xuất bản phẩm của Việt Nam được xuất bản trong nước: sách, kỹ yếu, tổng kết công trình nghiên cứu khoa học, tranh ảnh, bản nhạc, lịch,...
Cách trình bày	<ul style="list-style-type: none"> - Mã số (số thứ tự). Tên tác giả. Tên tài liệu (bằng ngôn ngữ gốc, tên tài liệu dịch sang tiếng quốc ngữ). - Nơi xuất bản, nhà xuất bản. Khổ. Ký hiệu xếp kho.
Phân loại	- Theo phân loại khoa học (BBK).
Bảng tra cứu/ chỉ dẫn	<ul style="list-style-type: none"> - Trong năm (Thư mục Quốc gia/ năm). + Bảng tra tên sách. + Bảng tra tên tác giả.

	<ul style="list-style-type: none"> + Bảng tra tên dịch giả. + Bảng tra cơ quan xuất bản, nhà xuất bản.
Phương pháp tra cứu tin	<ul style="list-style-type: none"> - Tra cứu tin hiện tại: <ul style="list-style-type: none"> + Theo lĩnh vực khoa học (Bảng phân loại BBK) trong thư mục của từng số. - Tra cứu tin hồi cố: <ul style="list-style-type: none"> + Theo các bảng tra trong Thư mục Quốc gia/ năm + Theo từng vấn đề trong mục lục. + Theo từng số lẻ của năm.

Các bản thư mục, tạp chí tóm tắt, mục lục dạng in... là công cụ giúp cho việc tra cứu thông tin hồi cố và hiện tại một cách nhanh chóng, thuận tiện và đầy đủ. Đặc biệt, tạp chí tóm tắt với phương pháp biên soạn, sắp xếp khoa học, có các công cụ làm việc thiết thực như:

- + Danh mục tên các tạp chí có bài được trích dẫn (Danh mục nguồn trích).

- + Tên viết tắt của các cơ quan tổ chức.

- + Từ viết tắt...

Tạp chí tóm tắt phản ánh thông tin kịp thời, đa dạng về tài liệu, đặc biệt là thông tin về văn bằng phát minh sáng chế giúp cho các nhà khoa học nghiên cứu không bị trùng lặp, tiết kiệm được công sức cũng như thời gian của họ. Tạp chí tóm tắt phản ánh đầy đủ những thành tựu khoa học kỹ thuật mới nhất trên thế giới, nhằm khắc phục những trở ngại trong quá trình thông tin bị phân tán và là công cụ hữu hiệu làm giảm thiểu những trở ngại do hàng rào ngôn ngữ tạo ra, trong nhiều trường hợp người dùng tin không phải tìm đọc tài liệu gốc.

So với hệ thống mục lục thư viện, các ấn phẩm thông tin và mục lục dạng in tiết kiệm thời gian tra tìm thông tin một cách đáng kể.

Ví dụ:

✓ Nêu tên sách và những bài báo của tác giả Võ Quý đăng trong các tạp chí từ năm 2000 đến 2004?

Nguồn tìm và phương pháp tìm:

1 - *Thư mục Quốc gia Việt Nam / Năm 2000 - 2004.*

. **Bảng tra: Tác giả** (Sắp xếp theo vần chữ cái Họ - Tên tác giả).

Vấn: V - Võ Quý

2- *Tạp chí tóm tắt tài liệu khoa học và công nghệ Việt Nam. Năm 2000 - 2004.*

. **Bảng tra: Tác giả** (Sắp xếp theo vần chữ cái Tên - Họ tác giả).

Vấn: Q - Võ Quý; Từ năm 2000: Theo bảng tra chung: tên tác giả / năm: 2000, 2001, 2002, 2003 và theo từng số lẻ năm 2004.

Ví dụ: Bảng tra Tác giả: Võ Quý năm 2002: 2002. 10/ 617, 2002.11/ 028 Cụ thể:

- 2002. 10/ 617. Võ Quý. Khảo cổ học Cao Bằng. - Khảo cổ học, 2001, N.2, 26 - 37.

- 2002. 11/ 028. Võ Quý. Về vấn đề quản lí vùng đệm ở Việt Nam, những khái niệm bước đầu. - Kỷ yếu hội thảo quốc tế vùng đệm các khu bảo tồn thiên nhiên Việt Nam, 2002, 42 - 49.

. Bảng tra: Tác giả Võ Quý năm 2003: 2003.04/ 704, 2003.06/ 647, 2003.08/ 033: Các tài liệu cụ thể:

- 2003.08/ 033. Võ Quý. Khảo cổ học Lâm Đông. - Khảo cổ học, 2003, N.2, 33- 47.

- 2003.06/ 647. Võ Quý. Tổng quan môi trường miền núi Việt Nam trong 20 năm qua: thực trạng và những vấn đề đặt ra. - Phát triển bền vững miền núi Việt Nam- 10 năm nhìn lại và những vấn đề đặt ra. - 2002, 98 - 115.

- 2003.04/ 704. Võ Quý. Cần phải làm gì để thúc đẩy công tác bảo tồn đa dạng sinh học ở Việt Nam?. - Thông tin những vấn đề sinh học ngày nay, 2002, vol. 8, N.4, 57 - 59.

Với các kết quả trên cùng tác giả Võ Quý nhưng nghiên cứu hai lĩnh vực hoàn toàn khác nhau. Trong trường hợp này, người tìm tin phải trực tiếp trao đổi với NDT để xác định họ cần tìm hiểu tài liệu của tác giả nào? Thuộc lĩnh vực khoa học nào? Trên cơ sở đó đáp ứng yêu cầu và còn có thể giúp họ tra cứu được những thông tin về lĩnh vực họ cần mà chưa cập nhật được.

* Cho biết tên các đề tài *luận án tiến sỹ về y học thực hành* bảo vệ từ năm 2000 - 2003.

Nguồn tìm và phương pháp tìm:

1- Tạp chí tóm tắt tài liệu khoa học và công nghệ Việt Nam. Năm 2000 – 2003.

. **Bảng tra:** Từ khoá mỗi số riêng biệt (Sắp xếp theo vần chữ cái từ khoá):

Vấn Y: *Y học thực hành*

. **Bảng để mục:** 76.29 *Y học thực hành (luận án)* (Theo từng số lẻ trong năm).

2 - Mục lục luận án tiến sỹ, tiến sỹ khoa học (Thư viện Quốc gia VN)

✓ Cho biết tên của *dịch giả* của tiểu thuyết *Cây sồi mùa đông*.

Nguồn tìm và phương pháp tìm:

1- Thư mục Quốc gia Việt Nam / năm (1983)

.*Bảng tra: Tên sách* (Sắp xếp theo vần chữ cái):

Vân C - Cây sồi mùa đông 1133

Dịch giả: Phạm Mạnh Hùng

.*Mục lục* (Sắp xếp theo BBK)

K31 - Tác phẩm văn học: Tiểu thuyết, truyện

Cây sồi mùa đông 1133

1133. Cây sồi mùa đông

Dịch giả: Phạm Mạnh Hùng

2 - *Mục lục chữ cái*: (Thư viện Quốc gia Việt Nam và nhiều thư viện khác).

Vân: C - Cây sồi mùa đông

Dịch giả: Phạm Mạnh Hùng

✓ Tìm các tài liệu có đề cập tới Ngôn ngữ học xuất bản năm 2002

Nguồn tìm và cách tra tìm

- *Thư mục Quốc gia Việt Nam* / năm 2002

Ngôn ngữ: J

- *Mục lục phân loại (BBK):*

Ngôn ngữ: J

Tra cứu Thư mục trích dẫn tạp chí khoa học.

Thư mục trích dẫn tạp chí khoa học (Science Citation Index) là một dạng sách có cấu trúc gồm các tài liệu trích dẫn đến một tập hợp tài liệu phản ánh về (hoặc) có liên quan đến một chủ đề xác định.

Đối tượng được phản ánh trong thư mục trích dẫn tạp chí khoa học là các tài liệu trích dẫn (Citing Documents) và các tài liệu được trích dẫn (Cited Documents).

Người tìm tin sử dụng chỉ dẫn chủ đề hoán vị. Trước tiên xác định chủ đề cần tìm trong thư mục trích dẫn tạp chí khoa học.

Có thể thực hiện một trong số các kỹ thuật tìm tin sau:

- Trường hợp đã biết một số tác giả trích dẫn hoặc được trích dẫn:

Từ tên tác giả được trích dẫn người tìm sử dụng thư mục trích dẫn và sau đó xác định được tên tài liệu được trích dẫn và trở lại thư mục nguồn để xác định các thông tin thư mục đầy đủ về danh mục tài liệu cần tra cứu.

- Trường hợp không xác định được tên tác giả cụ thể:

Người tìm tin có thể sử dụng trực tiếp (hoặc) thư mục nguồn hoặc thư mục trích dẫn để xác định được danh mục tài liệu để cập tới một chủ đề xác định (với điều kiện danh mục tài liệu đó được công bố trên các tạp chí xác định).

Tra cứu trong thư mục trích dẫn tạp chí khoa học ngày càng phát triển đáp ứng yêu cầu tin của cán bộ khoa học và công nghệ, các cán bộ quản lý khoa học, hoạt động sản xuất, thông tin khoa học công nghệ (bổ sung nguồn tin,...).

Như vậy là cùng với tạp chí tóm tắt, thư mục chuyên đề, thư mục chuyên ngành, thư mục trích dẫn tạp chí khoa học giúp cho người

dùng tin kiểm soát được một cách đầy đủ, toàn diện tới nguồn tài liệu phản ánh về/ hoặc có liên quan với một chủ đề xác định (hạn chế những trở ngại của quá trình phân tán thông tin gây nên) và điều đó không phụ thuộc vào các tài liệu đó được lưu giữ ở thư viện nào, thể hiện bằng ngôn ngữ nào (hạn chế được trở ngại do hàng rào ngôn ngữ tạo nên).

2. TRA CỨU SỐ LIỆU VÀ DỮ KIỆN

Tra cứu thông tin dữ kiện (số liệu, dữ kiện) là một dạng tra cứu rất thông dụng tại các cơ quan thông tin, thư viện, đặc biệt trong nền kinh tế thị trường và khoa học - công nghệ phát triển. Tra cứu thông tin dữ kiện là quá trình tách ra khỏi nguồn tìm những số liệu, dữ kiện liên quan tới một đối tượng nào đó như: Số liệu thống kê, tính chất của vật, vật liệu, tính năng kỹ thuật của máy móc thiết bị,...

Hệ thống phiếu tra cứu dữ kiện thường được xây dựng ở các cơ quan thông tin thư viện chuyên ngành, các hệ thống thông tin hành chính, các cơ sở sản xuất - kinh doanh.

2.1. Nguồn tra cứu chính

- Bộ phiếu tra cứu dữ kiện:

Bộ phiếu tra cứu về vật liệu,

Bộ phiếu tra cứu về máy móc, thiết bị,

Bộ phiếu tra cứu về quy trình công nghệ, ...

- Tài liệu tra cứu:

Bách khoa thư,

Từ điển,
Sách tra cứu chuyên ngành,
Almanach,
Tập bản đồ (Atlas),
Sách tra cứu địa lí (Gazetteers),
Niên giám,...

Ngoài các nguồn tra cứu thông dụng trên, thông tin về các dữ kiện còn có thể tra cứu trong một số nguồn khác như: nguồn tra cứu về các tổ chức chính trị - xã hội, nguồn tra cứu về tiểu sử - nhân vật, nguồn tra cứu về các ấn phẩm công...

2.2. Phương pháp tra cứu

Khác với tra cứu thông tin thư mục, tra cứu thông tin dữ kiện là việc tra cứu các số liệu, dữ kiện có trong bản thân tài liệu như: Các số liệu thống kê, thời gian, công thức hoá học, tính chất của vật liệu, thông số kỹ thuật của máy móc thiết bị, khái niệm khoa học,...

Ví dụ:

- ✓ Vua Lê Thánh Tông lên ngôi năm nào? Trữ trị bao nhiêu năm?
- ✓ Ngôi sao bóng đá Maradona đã từng tham gia những câu lạc bộ bóng đá nào ở châu Âu?
- ✓ Thủ tướng Trung Quốc hiện nay là ai? Vài nét về tiểu sử?
- ✓ Năm 2003 Việt Nam xuất khẩu bao nhiêu tấn gạo?
- ✓ Cho biết công thức hàm (toán học) exponenta?
- ✓ Giải thích thuật ngữ Engim,...

Tại các cơ quan thông tin thư viện, các cơ quan thực hiện dịch vụ hỏi - đáp, tra cứu, tư vấn thông tin, thông tin dữ kiện được sử dụng để phục vụ chủ yếu cho những yêu cầu tin trong các lĩnh vực sản xuất, kinh doanh, nghiên cứu khoa học - công nghệ, quản lý hành chính. Để tìm kiếm thông tin dữ kiện việc đầu tiên là xác định được các đối tượng cần tìm và các thuộc tính đối tượng phản ánh.

Ví dụ:

- + Các nguồn tài nguyên thiên nhiên,
- + Các chủng loại hàng hoá, sản phẩm công, nông, lâm, ngư nghiệp,
- + Tên các hãng, công ty: khả năng tài chính, lĩnh vực đầu tư sản phẩm,
- + Mẫu mã, nhãn hiệu hàng hoá,...

Cũng như tra cứu thông tin thư mục, tra cứu thông tin dữ kiện cũng phải dựa vào các yêu cầu tin (câu hỏi) cụ thể để lựa chọn nguồn tra cứu. Thông thường chúng ta hay gặp 5 loại câu hỏi cơ bản và xác định với câu hỏi nào sẽ sử dụng loại tài liệu tra cứu nào cho phù hợp đó là:

- Cái gì?
- Ai?
- Ở đâu?
- Khi nào?
- Tại sao?

Có thể tham khảo ở trong bảng 5.

Bảng 4. Các câu hỏi thường sử dụng

Câu hỏi		Mục đích	Nguồn tìm
Cái gì? (What?)	Các từ (Words)	. Ý nghĩa của từ . Cách phát âm . Từ đồng nghĩa, trái nghĩa . Cách sử dụng	- Từ điển - Từ điển từ chuẩn (Thesauri)
	Sự vật (Things)	Khái niệm, kiến thức (Background)	- Bách khoa thư
. Sự kiện và con số (Facts and figures)		- Almanac - Sách tra cứu	
Ai? (Who?)	Nhân vật (People)	Tiểu sử (Background)	- Từ điển danh nhân - Bách khoa thư
		Địa phương (Location)	- Từ điển / danh nhân địa phương
Tại sao? (Why?)	Thông tin (Information)	Sự kiện cập nhật (Current events) ----- Sự kiện và con số (Facts and figures) ----- Khái niệm, kiến thức (Background).	- Bảng tra ----- - Từ điển, - Sách tra cứu, - Niên giám thống kê ----- - Bách khoa thư

Ở đâu? (Where?)	Địa danh (Places)	Lãnh thổ: vùng, miền, tỉnh, thành phố (Location) ----- Sự kiện và con số (Facts and figures) ----- Kiến thức - hiểu biết (Lịch sử và sự kiện)	- Bản đồ - Tập bản đồ (Atlas) - Sách tra cứu địa lý (gazetteer) ----- - Almanach - Sách tra cứu ----- - Bách khoa thư - Bản đồ chuyên môn, bản đồ chuyên ngành - Từ điển địa danh
Khi nào? (When ?)	. Sự kiện và ngày tháng (Events and dates)	Sự kiện và con số (Facts and figures) . Lịch sử ----- Khái niệm, kiến thức (Background).	- Almanach và niên giám ----- - Bách khoa thư

Bất cứ câu hỏi nào cũng phải biết phân tích vấn đề, trên cơ sở đó lựa chọn công cụ tra cứu phù hợp.

Ví dụ:

* Cho biết thời gian và diễn biến cuộc khởi nghĩa Thái Nguyên?
(Câu hỏi: Khi nào?)

Nguồn tìm:

Để làm rõ thời gian và diễn biến sự kiện, ta phải tìm ở Từ điển bách khoa Việt Nam, Almanach, Niên giám,...

Từ điển bách khoa Việt Nam. T.2: Văn K - Khởi nghĩa Thái Nguyên:

. Biên niên sử.

* Cho biết đỉnh Everet cao bao nhiêu mét? Ai là người đầu tiên chinh phục nó? (Trả lời câu hỏi: Cái gì, và ai).

Nguồn tìm:

Bách khoa thư (các nước).

Từ điển bách khoa Việt Nam:

Từ điển bách khoa Việt Nam. T.2. Văn E - Everest: Xem: Chômôlungma.

Chômôlungma cao: 8880 m.

* Nhà văn L. Tolstoi mất năm nào? (Trả lời câu hỏi Ai).

Nguồn tìm:

. Bách khoa thư (Các nước).

. Almanach. Những nền văn minh thế giới

. Từ điển danh nhân thế giới,...

* Tại sao Mỹ đánh Iraq (Lần thứ nhất)?

Trả lời câu hỏi Tại sao?

Nguồn tìm:

. Bách khoa thư (Các nước).

. Niên giám.

. Almanach.

. Từ điển

* Sông Danip chảy qua những nước nào?

Trả lời câu hỏi: Cái gì?

Nguồn tìm:

. Bách khoa thư

. Từ điển địa danh

* Cho biết Nam Phi nằm ở vĩ độ và kinh độ nào?

Trả lời câu hỏi: Ở đâu?

Nguồn tìm:

. Bách khoa thư,

. Từ điển địa danh,

. Sổ tay tra cứu địa danh,

. Bản đồ địa lý,

. Tập bản đồ,

3. Một số vấn đề cần lưu ý trong quá trình tra cứu

Với sự phát triển mạnh mẽ của khoa học và công nghệ trong những thập niên vừa qua, dẫn tới thông tin khoa học - công nghệ là một trong những nguồn tin rất chóng bị lỗi thời trong đó không loại trừ nhiều tài liệu tra cứu có liên quan. Vì vậy, khi sử dụng tài liệu tra cứu cần đánh giá xem tài liệu cụ thể đó có còn phù hợp với yêu cầu tin hay không? Nghĩa là tài liệu có cung cấp thông tin cập nhật, đảm bảo độ tin cậy, đầy đủ,...

Để kiểm tra vấn đề đó nên xem xét tới một số yếu tố cụ thể có liên quan sau:

- Tác giả,

- Mục đích xuất bản của tài liệu,
- Phạm vi bao quát đề tài,
- Quan điểm,
- Cách sắp xếp,
- Chất lượng của các bảng tra,
- Cách trình bày,
- Tính cập nhật thông tin,...

*** Tác giả:**

Nhà xuất bản có danh tiếng có ý nghĩa quan trọng và có thể sử dụng như một tiêu chuẩn để đánh giá chất lượng của tài liệu. Lời tựa (mở đầu) thường liệt kê tên những người chủ biên, ban biên tập, cố vấn, cộng tác viên và cả trình độ, học hàm, học vị của họ,... Những thông tin chính xác đó có ý nghĩa và tác động đến yếu tố tác giả, dựa vào đó có thể so sánh lần xuất bản sau so với những lần trước; tài liệu của tác giả này với tác giả khác có cùng tên sách.

*** Mục đích xuất bản:**

Đa phần tài liệu tra cứu trong phần mở đầu đều có phần giải thích mục đích (lí do) xuất bản tài liệu và đối tượng phục vụ của tài liệu.

*** Phạm vi:**

Khi cần tra cứu phải xem xét tài liệu có bao hàm lĩnh vực mình lựa chọn hay không? Có những nguồn tra cứu chi dành riêng cho các chuyên gia, còn một số khác dùng cho các đối tượng phổ thông. Tuy nhiên, tác giả hay nhà xuất bản có thể cũng cần thoả hiệp giữa sự uyên bác và đại chúng trong khi xuất bản, trên cơ sở đó định giá cả cho sách.

*** Quan điểm:**

Theo suy nghĩ của nhiều người, bất cứ tài liệu tra cứu nào của các quốc gia cũng có thể tra tìm và trả lời câu hỏi một cách chính xác, đầy đủ cho đúng đối tượng người dùng tin, tuy nhiên trong thực tế không ít tài liệu tra cứu mang quan điểm của một/ một số quốc gia cụ thể - đó là nhu cầu thông tin tới mỗi quốc gia riêng và do đó nó có sự khác biệt với quan điểm của nhiều nước. Thông thường khi xem xét tài liệu tra cứu cần biết nó xuất bản ở đâu? Quan điểm của nước đó có ảnh hưởng tới thông tin mình cần tra tìm không?

*** Cách sắp xếp:**

Đa số người dùng tin muốn tra cứu thông tin nhanh chóng nên họ ưa thích nguồn tin đơn giản, dễ tham khảo và đa số tài liệu tra cứu dễ sử dụng do cách sắp xếp khoa học. Tuy nhiên, mỗi nguồn tra cứu có mục đích khác nhau nên có cách sắp xếp khác nhau. Ví dụ: Các bách khoa thư, từ điển ngôn ngữ, từ điển nhân vật hay từ điển thuật ngữ thường được sắp xếp theo hình thức: nghĩa là theo vần chữ cái của ngôn ngữ tài liệu, còn các loại sách tra cứu chuyên ngành đa phần được sắp xếp theo nội dung chi tiết, cụ thể của chuyên ngành đó.

*** Chất lượng của các bảng tra.**

Các bảng tra trong tài liệu tra cứu phải bao quát được nội dung phản ánh của tài liệu, giúp cho việc sử dụng ấn phẩm được dễ dàng và nhanh chóng. Nó có thể bao gồm các bảng chỉ dẫn / bảng tra cứu trực tiếp tới đối tượng cần tìm cũng như các bảng tra chéo, hoặc các chỉ chỗ: "Xem", "Cũng xem", ... Nó cũng có thể rất đơn giản chỉ có mục lục của tài liệu. Khi đánh giá cần xem xét cụ thể.

*** Cách trình bày của tài liệu:**

Khi xem xét một tài liệu tra cứu cụ thể, cần quan tâm tới đặc trưng loại sách, hình thức xuất bản, tranh ảnh, bản đồ, chất lượng giấy, màu sắc, dạng bìa sách, cách sắp xếp chúng... Tất cả những yếu tố đó cũng sẽ ảnh hưởng tới chất lượng của tài liệu, nó giúp chúng ta quyết định lựa chọn nguồn tra cứu này mà không chọn những nguồn tra cứu khác. Có tài liệu tra cứu được trang trí nhằm thu hút sự chú ý của người đọc: Màu sắc, tranh ảnh, bản đồ và phông chữ đặc biệt,... những cái đó cũng quan trọng và góp phần làm tăng giá trị của tài liệu, tuy nhiên nó không thể làm lạc hướng việc giới thiệu và đánh giá chuẩn xác thông tin cụ thể mà tài liệu đó phản ánh.

*** Tính cập nhật thông tin:**

Tính cập nhật thông tin là một yếu tố rất quan trọng của một ấn phẩm tra cứu. Nó được thể hiện trong Lời nói đầu/ Lời giới thiệu. Thông thường, khi tra cứu bao giờ người sử dụng cũng quan tâm tới những tài liệu mới xuất bản, song khi tra cứu cần phải lưu ý rằng năm xuất bản không phải bao giờ cũng thể hiện tính cập nhật nội dung thông tin vì có nguồn tra cứu lấy năm biên soạn là thời gian xuất bản tài liệu và nhiều ấn phẩm xuất bản kéo dài nhiều năm (có khi từ 5 - 10 năm), ví dụ như các loại bách khoa thư: Encyclopedia Britanica, Grande Larouse, Encyclopedia Americana,... Đối với những loại tài liệu đó ngoài những tập chính cần quan tâm tới các niên giám (yearbooks) hay danh mục tra cứu (supplementary lists, Index), chúng xuất bản định kỳ hàng năm, thông tin được thường xuyên bổ sung và cập nhật.

Trong các nguồn tra cứu chủ yếu trên có một số nguồn tra cứu thông tin cũng rất chóng lỗi thời như: tài liệu tra cứu về các tổ chức -

cơ quan, các loại niên giám thống kê, niên biểu, đặc biệt là sách tra cứu về địa lí có một số lưu ý khi sử dụng:

Tính chính xác của thông tin trong tài liệu tra cứu: song song với thời gian xuất bản của tài liệu cộng với sự thay đổi về biên giới quốc gia, tên gọi và lãnh thổ của một số nước cũng góp phần làm cho thông tin không phải tài liệu nào cũng có khả năng cập nhật thường xuyên,

Quan điểm thống nhất về tên địa danh: nó có thể là vấn đề chuyển mãu chữ không phải chữ Latinh sang dạng khổ chữ khác (chữ Ả-rập, chữ tượng hình,...), đa số từ điển tuân thủ và công nhận theo chuẩn, tuy nhiên không ít từ điển không tuân thủ theo quy định (trong đó có một số từ điển xuất bản ở Việt Nam những năm 1960 - 1980).

Tính cân đối: Tập bản đồ thế giới của nhiều nước xuất bản đa phần tuân thủ tính cân đối về tỷ lệ, song có thể có những nước rất nhấn mạnh tầm quan trọng của nước mình và nó phá vỡ tính cân đối trong ấn phẩm.

Chương 4

Tìm tin tự động hoá

Tìm tin bằng phương pháp truyền thống được tiến hành từ rất lâu trước khi có máy tính điện tử. Sự ra đời của máy tính điện tử và ứng dụng chúng vào hoạt động thông tin thư viện từ đầu những năm 50 của thế kỷ XX đã dẫn đến sự xuất hiện của các cơ sở dữ liệu và ngân hàng dữ liệu. ứng dụng máy tính điện tử đã tạo ra những bước phát triển mạnh mẽ trong các hoạt động thông tin thư viện, dẫn đến quá trình tin học hoá nhiều công đoạn của hoạt động thông tin thư viện. Nhờ ứng dụng máy tính điện tử, các chuyên gia đã sáng tạo ra những hệ thống tìm tin dựa trên cơ sở máy tính điện tử. Nhờ sự phát triển mạnh mẽ của công nghệ thông tin, những hệ thống tìm tin tự động hoá cũng ngày càng được phát triển và hoàn thiện. Tìm tin trong các hệ thống tìm tin tự động hoá đã trở nên một công việc phổ biến và bình thường tới mức trong nhiều tài liệu hiện đại thuật ngữ tìm tin (information retrieval) chủ yếu dùng để chỉ việc tìm tin có sử dụng máy tính điện tử, còn việc tra tìm thông tin theo phương pháp truyền thống dựa trên các công cụ truyền thống được gọi là công tác tra cứu (reference work).

Với sự phát triển nhanh đến chóng mặt và ứng dụng rộng rãi của công nghệ thông tin và viễn thông vào hoạt động thông tin thư viện,

số lượng các cơ sở dữ liệu, ngân hàng dữ liệu, dịch vụ thông tin trực tuyến và các sản phẩm thông tin trên CD-ROM tăng lên không ngừng tạo ra nguồn tin điện tử khổng lồ với những dạng thức đa dạng, phong phú. Để khai thác hiệu quả nguồn tin nói trên đòi hỏi người tìm tin phải có những kiến thức và kỹ năng tìm tin tự động hoá. Trong phần này để cập đến hệ thống tìm tin tự động hoá, phương pháp tìm tin tự động hoá, quá trình và chiến lược tìm tin tự động hoá.

I. DẠNG TÌM TIN TỰ ĐỘNG HOÁ

Việc phân loại các dạng tìm tin tự động hoá có tính chất tương đối. Chúng ta có thể phân loại tìm tin tự động hoá dựa trên:

- Nơi lưu giữ cơ sở dữ liệu;
- Theo người thực hiện quá trình tìm.

Dựa vào nơi lưu giữ cơ sở dữ liệu để truy cập, người ta có thể chia tìm tin tự động hoá thành các dạng:

- Tìm tin cục bộ,
- Tìm tin trực tuyến,
- Tìm tin trên CD-ROM.

Tìm tin cục bộ là quá trình tìm tin được thực hiện với cơ sở dữ liệu được lưu trữ ngay trên bộ nhớ của chính máy tính điện tử mà người tìm tin đang sử dụng. Tìm tin cục bộ cũng có thể được gọi là tìm tin trên máy tính điện tử. Trong trường hợp này, phần mềm lưu trữ và tìm tin tương ứng được cài đặt cùng với cơ sở dữ liệu cục bộ.

Tìm tin trực tuyến (thuật ngữ tiếng Anh là On-line Information Retrieval) là thuật ngữ chỉ việc tìm thông tin được lưu giữ trong cơ sở

dữ liệu trên máy tính từ xa. Quá trình tìm tin được gọi là trực tuyến (on-line) được đặc trưng bởi việc truy cập đến các cơ sở dữ liệu từ xa và người tìm tin phải sử dụng mạng máy tính, mạng viễn thông, những phần mềm chuyên dụng. Tìm tin trực tuyến thường được thực hiện thông qua các tổ chức kinh doanh dịch vụ thông tin trực tuyến.

Tìm tin trên CD-ROM là hoạt động tìm kiếm thông tin được tiến hành với các cơ sở dữ liệu lưu giữ trên CD-ROM. Nhờ sự phát triển của công nghệ sản xuất CD-ROM có khả năng lưu trữ lớn, nhiều nhà sản xuất cơ sở dữ liệu và các dịch vụ thông tin trực tuyến đã đưa nhiều loại cơ sở dữ liệu lên đĩa CD-ROM và phổ biến chúng tới người sử dụng trực tiếp. Những đĩa CD-ROM này có thể chứa phần mềm tìm tin chuyên dụng, cơ sở dữ liệu, tài liệu toàn văn. Với các cơ sở dữ liệu trên CD-ROM, người tìm tin không cần phải thực hiện tìm tin trực tuyến tới các cơ sở dữ liệu từ xa. Tìm tin trên CD-ROM đôi khi còn được gọi là tìm tin tại chỗ (on-site information retrieval) bởi các cơ sở dữ liệu truy cập trực tuyến trước đây, do được đưa lên CD-ROM nên tìm kiếm được theo chế độ cục bộ.

Trong nhiều trường hợp, phần mềm tìm tin có thể được cài đặt vào ổ đĩa cứng còn cơ sở dữ liệu thì vẫn được lưu trên CD-ROM. Khi đó người ta truy cập chương trình tìm trên đĩa cứng. Trong trường hợp khác, một phần dữ liệu được cài đặt trên đĩa cứng của máy tính điện tử, một phần khác vẫn được lưu trên CD-ROM. Ví dụ bộ đĩa Thư viện điện tử IEL (IEEE/IEE Electronic Library) có chương trình tìm kiếm và cơ sở dữ liệu được cài trên đĩa cứng của máy tính trong khi dữ liệu phần toàn văn được đặt trên các đĩa CD-ROM. Do đó khi khai thác IEL phải sử dụng cả cơ sở dữ liệu trên đĩa cứng đồng thời sử dụng cả các đĩa CD-ROM chứa thông tin toàn văn.

Tìm tin trên CD-ROM có thể được thực hiện theo hai phương thức:

- Tìm tin theo chế độ độc lập (stand-alone).
- Tìm tin theo chế độ mạng.

Việc tìm tin trên CD-ROM được tiến hành theo chế độ mạng khi các cơ sở dữ liệu trên CD-ROM được đưa lên mạng máy tính bằng các thiết bị đọc CD-ROM được nối mạng như tháp CD-ROM (CD Tower) hoặc hộp CD-ROM (Jukebox).

Dựa vào người tìm tin, người ta có thể chia tìm tin tự động hoá thành:

- Tìm tin trực tiếp;
- Tìm tin gián tiếp.

2. CÚ PHÁP CỦA TÌM TIN TỰ ĐỘNG HOÁ

Tìm tin tự động hoá gắn liền với hệ thống tìm tin tự động hoá. Cú pháp của tìm tin tự động hoá, cho dù đó là tìm tin trên hệ thống tìm tin trực tuyến hay trên CD-ROM hoặc trong cơ sở dữ liệu trên máy tính cá nhân, là tương tự như nhau. Việc nắm bắt được những nguyên tắc cơ bản khi xây dựng biểu thức tìm đảm bảo cho người tìm tin khai thác hiệu quả hệ thống.

Trong tìm tin tự động hoá, chúng ta phải nắm vững những khái niệm hết sức quan trọng như biểu thức tìm, thuật ngữ tìm và các toán tử. Cú pháp tìm tin được sử dụng để tạo ra các biểu thức tìm (còn gọi là phương trình tìm) bằng các thuật ngữ tìm và việc kết hợp các thuật ngữ bằng những toán tử khác nhau.

2.1. Biểu thức tìm

Trong tìm tin tự động hoá, một bước quan trọng là xây dựng một biểu thức tìm. *Biểu thức tìm là một phương trình tìm trong đó các thuật ngữ tìm tin được liên kết với nhau bằng các toán tử thể hiện các quan hệ giữa những thuật ngữ tìm do người tìm tin xác lập phù hợp với cú pháp của hệ thống.*

Biểu thức tìm có thể có các thành phần như thuật ngữ tìm, toán tử, các ký hiệu chặt cắt, các dấu đóng ngoặc và mở ngoặc đơn, những ký hiệu tìm kiếm mở rộng hoặc hạn chế...

Thuật ngữ tìm tin (search term) là từ, những từ, cụm từ, con số được lựa chọn thể hiện yêu cầu tin. Thuật ngữ tìm tin có thể là một từ trong nhan đề, từ khoá, từ chuẩn, chỉ số phân loại, tên tác giả, tên nhà xuất bản, năm xuất bản. Ví dụ:

Nguyễn Tuân

lúa

nông nghiệp

87.01

1990

Thuật ngữ tìm tin không chỉ là từ, cụm từ, con số mà còn có thể là ký hiệu mà hệ thống tìm tin sử dụng để thể hiện những biểu thức tìm đã thực hiện trước đó. Chẳng hạn, chương trình CDS/ISIS sử dụng ký hiệu # và con số để thể hiện biểu thức tìm đã thực hiện (Ví dụ #2), còn chương trình DIALOG sử dụng ký tự S và con số kèm theo (Ví dụ S3). Ví dụ về biểu thức tìm với thuật ngữ tìm có kết hợp biểu thức tìm trước đó (với CDS/ISIS):

#2 * chọn giống

trong đó:

#2 là thể hiện biểu thức tìm số 2

hoặc biểu thức tìm với chương trình tìm tìm của Dialog:

S1 AND Computers

trong đó:

S1 là thể hiện biểu thức tìm số 1

Nói tóm lại bất cứ một từ hay ký hiệu nào được nhập vào mà không phải những ký hiệu đã được hệ thống sử dụng làm toán tử, làm các ký hiệu mở rộng/thu hẹp, toán hạng, so sánh điều có thể được hệ thống tìm tin hiểu là thuật ngữ tìm.

Trong tìm tin tự động hoá, có thể phân biệt hai loại thuật ngữ tìm là:

- Thuật ngữ chính xác.
- Thuật ngữ chặt chẽ.

Thuật ngữ chính xác (*precise term*) là phần tử tìm kiếm được xác định đối với hệ thống tìm tin. Ta phải hiểu và nắm chính xác các thuật ngữ này của hệ thống. Khi sử dụng thuật ngữ chính xác, hệ thống tìm tin sẽ thực hiện các phép so sánh chúng với những thuật ngữ đã được lưu trong tệp đảo hoặc tệp chỉ số của cơ sở dữ liệu. Những biểu ghi có chứa thuật ngữ trùng khớp với thuật ngữ tìm tin được chúng ta nhập vào biểu thức tìm sẽ được coi là phù hợp và sẽ được hệ thống chọn ra. Mọi sai khác, dù là nhỏ nhất, giữa thuật ngữ nhập vào biểu thức và thuật ngữ có trong tệp đảo, đều coi là không trùng hợp. Kết quả là những biểu ghi chứa các thuật ngữ có sai khác sẽ không thể được hệ

thống chọn ra. Ví dụ, nếu ta nhập thuật ngữ "hoá học" (dấu sắc ở chữ A) vào biểu thức tìm trong khi các biểu ghi trong cơ sở dữ liệu lại được mô tả bằng thuật ngữ "hóa học" (dấu sắc ở chữ O) thì kết quả sẽ là không có biểu ghi phù hợp (không tìm ra) vì với hệ thống tìm tin "hoá học" khác "hóa học".

Một loại hình thuật ngữ tìm tin đặc biệt gọi là **thuật ngữ chặt cụt** (*truncated term*) được nhiều hệ thống tìm tin sử dụng. *Thuật ngữ chặt cụt là thuật ngữ tìm có chứa phần gốc của thuật ngữ và ký hiệu chặt cụt.*

Ví dụ với chương trình CDS/ISIS, thuật ngữ chặt cụt có thể như sau:

nông nghiệp\$

lúa\$

fish\$

Trong những ví dụ trên, các cụm từ "nông nghiệp", "lúa" hoặc "fish" là phần gốc của thuật ngữ, còn ký hiệu "\$" là ký hiệu chặt cụt theo quy định của chương trình CDS/ISIS. Kỹ thuật tìm tin sử dụng thuật ngữ chặt cụt là kỹ thuật rất quan trọng và được sử dụng khá thường xuyên trong tìm tin tự động hoá (xem chi tiết phần phép tìm tin chặt cụt).

2.2. Toán tử

Biểu thức tìm tin có thể rất đơn giản chỉ gồm một từ, một cụm từ nhưng cũng có thể phức tạp với nhiều thuật ngữ tìm tin. Để kết hợp được các thuật ngữ tìm tin phải sử dụng những toán tử. Toán tử là những loại lệnh đặc biệt được hệ thống tìm tin quy định để thể hiện những quan hệ giữa các thuật ngữ tìm tin. Đại đa số các hệ thống tìm tin đều cho phép xây dựng những biểu thức tìm phức tạp sử dụng các toán tử.

Để tìm tin hiệu quả, người ta sử dụng và phối hợp những phương pháp khác nhau trong tạo lập biểu thức tìm. Trong tìm tin tự động hoá, người ta có thể áp dụng một số phương thức tìm tin như:

- Sử dụng các toán tử boole.
- Sử dụng toán tử lân cận.
- Sử dụng phép tìm chặt cụt (truncation).
- Sử dụng phép tìm giới hạn theo trường.
- Sử dụng phép tìm so sánh.

Người tìm tin có thể sử dụng một phương thức hoặc có thể phối hợp nhiều phương thức trong cùng một biểu thức tìm. Sau đây sẽ trình bày chi tiết những phương thức tạo lập biểu thức tìm cơ bản.

2.3. Sử dụng toán tử boole

Trong biểu thức tìm, người ta có thể sử dụng một từ, một cụm từ làm thuật ngữ tìm tin. Tuy nhiên trong thực tế, với yêu cầu tìm tin phức tạp, không thể sử dụng một thuật ngữ duy nhất mà phải cần nhiều thuật ngữ tìm để thể hiện yêu cầu tìm tin nhằm đạt được kết quả đáp ứng được yêu cầu của người dùng tin. Trong trường hợp đó, người tìm tin sử dụng toán tử để thể hiện các quan hệ giữa những thuật ngữ tìm tin. Các toán tử boole chính là công cụ như vậy.

Các toán tử boole (tiếng Anh là Boolean operators), được đặt tên theo tên nhà toán học người Anh George Boole (1815-1864), người đã đưa ra khái niệm toán tử khi thể hiện sự kết hợp các mệnh đề theo hình thức logic. George Boole đưa ra 3 toán tử là:

- Toán tử giao (phép nhân, logical product), thường được ký hiệu là * hoặc AND;

- Toán tử cộng logic (phép cộng, logic sum), thường được ký hiệu là + hoặc OR (hoặc);

- Toán tử trừ logic (phép trừ, logical difference), thường được ký hiệu là - hoặc NOT.

Những toán tử do G.Boole đề xuất thường được ký hiệu bằng ký hiệu toán tử **AND**, **OR** và **NOT**. Sau này, nhà toán học John Venn đã thể hiện những quan hệ logic của toán tử boole dưới dạng biểu đồ, thường được gọi là biểu đồ Venn.

Toán tử Boole là dạng toán tử được sử dụng thường xuyên nhất trong quá trình tìm tin. Các toán tử này cho phép thể hiện ba loại quan hệ cơ bản giữa các thuật ngữ trong biểu thức tìm: quan hệ giao, quan hệ hợp và quan hệ loại trừ. Thông thường, các hệ thống tìm tin tự động hoá sử dụng các ký hiệu AND, OR và NOT để thể hiện các toán tử boole. Tuy nhiên, cũng có những hệ thống tìm tin sử dụng những ký hiệu khác cho các toán tử boole. Ví dụ, chương trình CDS/ISIS sử dụng ký hiệu * (dấu sao) cho toán tử AND, dấu + (dấu cộng) cho toán tử OR, dấu ^ (dấu mũ) cho toán tử NOT. (Để giới thiệu chung, chúng ta sử dụng ký hiệu AND, OR và NOT). Khi áp dụng cho hệ thống cụ thể, chúng ta phải tìm hiểu hệ thống và sử dụng ký hiệu do hệ thống đưa ra.

2.3.1. Toán tử AND

Toán tử AND tạo ra quan hệ giao và được sử dụng để thể hiện yêu cầu phải có sự xuất hiện đồng thời của những thuật ngữ hoặc khái niệm được dùng để định chỉ cho biểu ghi.

Nếu viết biểu thức tìm có dạng tổng quát:

A AND B

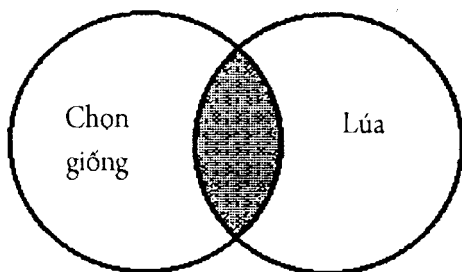
trong đó A và B là hai thuật ngữ tìm, AND là toán tử, thì toán tử AND yêu cầu các biểu ghi trong tập hợp tìm được phải được định chỉ số đồng thời với hai thuật ngữ A và B. Các biểu ghi được định chỉ số là A và B sẽ là những biểu ghi phù hợp. Toán tử AND có tác dụng là chính xác hoá yêu cầu tìm kiếm, thu hẹp kết quả tìm tin.

Ví dụ, nếu ta muốn tìm thông tin về "chọn giống lúa", chúng ta có thể tạo một biểu thức tìm có sử dụng toán tử AND nối hai thuật ngữ tìm tin như sau:

lúa AND chọn giống

Những biểu ghi được định chỉ số là chọn giống và lúa sẽ là biểu ghi phù hợp. Như vậy, tập hợp biểu ghi (tài liệu) thoả mãn yêu cầu tìm chính là giao của hai tập hợp: tập hợp biểu ghi thoả mãn yêu cầu có thuật ngữ "lúa" và tập hợp thoả mãn yêu cầu có thuật ngữ "chọn giống".

Biểu thức tìm sử dụng toán tử AND như trên có thể được trình bày bằng biểu đồ Venn như trong hình 5.



Hình 5. Biểu đồ Venn về tác động của toán tử AND

Trong hình 5, vòng tròn bên phải là tập hợp những biểu ghi được định chỉ số với thuật ngữ "lúa", tập hợp bên trái là tập hợp những biểu ghi được định chỉ số với thuật ngữ "chọn giống" và phần giao nhau (tô đậm) là tập hợp những biểu ghi thoả mãn yêu cầu của biểu thức tìm. Đó là những biểu ghi được định chỉ số đồng thời với cả hai thuật ngữ **Chọn giống và Lúa.**

2.3.2. Toán tử OR

Toán tử OR sẽ tạo ra quan hệ hợp và được sử dụng để thể hiện yêu cầu rằng biểu ghi thoả mãn yêu cầu là biểu ghi chỉ cần chứa một trong những thuật ngữ trong biểu thức tìm.

Nếu viết biểu thức tìm có dạng:

A OR B

trong đó A và B là hai thuật ngữ tìm, OR là toán tử, thì toán tử OR sẽ tìm ra các biểu ghi định chỉ số với một trong hai thuật ngữ hoặc cả hai thuật A và B. Tất cả biểu ghi được định chỉ số với A hoặc B hoặc cả hai đều là phù hợp với yêu cầu tìm.

Toán tử OR có tác dụng mở rộng kết quả tìm tin, nâng cao độ đầy đủ.

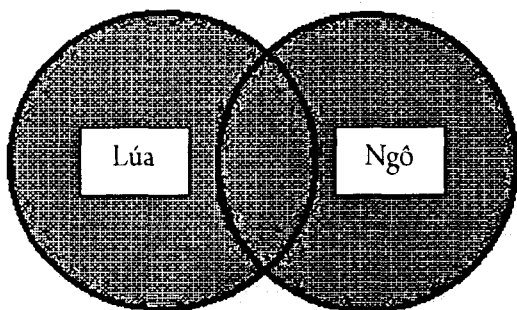
Ví dụ, nếu ta muốn tìm tài liệu về ngô hoặc lúa, ta có thể sử dụng toán tử OR để tạo ra biểu thức tìm như sau:

ngô OR lúa

Những biểu ghi được định chỉ số là "lúa" hoặc "ngô" sẽ đều là biểu ghi phù hợp. Như vậy tập hợp biểu ghi (tài liệu) thoả mãn yêu cầu tìm chính là hợp của hai tập hợp: tập hợp biểu ghi thoả mãn yêu cầu có thuật ngữ "lúa" và tập hợp thoả mãn yêu cầu có thuật ngữ "ngô".

Biểu thức Tác dụng của toán tử OR có thể được trình bày bằng sơ đồ Venn như hình 6:

Trong hình 6, vòng tròn bên phải là tập hợp biểu ghi được định chỉ số với thuật ngữ "**ngô**", vòng tròn bên trái là tập hợp biểu ghi được định chỉ số với thuật ngữ "**lúa**". Toàn bộ phần tô đậm là tập hợp các biểu ghi thoả mãn yêu cầu của biểu thức tìm. Đó là những biểu ghi được định chỉ số hoặc với thuật ngữ "**lúa**" hoặc với "**ngô**" hoặc với cả hai.



Hình 6. Sơ đồ Venn về tác động của toán tử OR

2.3.3. Toán tử NOT

Toán tử NOT sẽ tạo ra quan hệ loại trừ và được sử dụng để yêu cầu rằng biểu ghi tìm được phải là biểu ghi được định chỉ số bằng thuật ngữ ở phía trước toán tử và không được định chỉ số ở phía sau của toán tử.

Nếu viết biểu thức tìm có dạng:

A NOT B

trong đó A và B là hai thuật ngữ tìm, NOT là toán tử, thì toán tử NOT sẽ tìm ra các biểu ghi định chỉ số với thuật ngữ A, nhưng không được định chỉ số với thuật ngữ B.

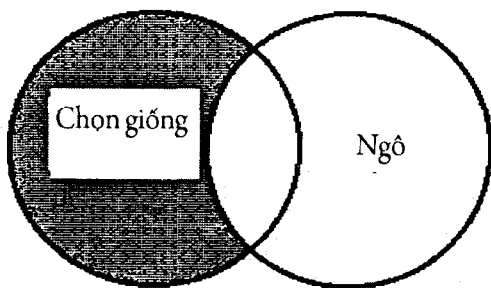
Khác với hai trường hợp toán tử AND và OR, trong trường hợp sử dụng toán tử NOT, thứ tự thuật ngữ trong biểu thức tìm có ý nghĩa quan trọng. Kết quả của biểu thức tìm "A NOT B" sẽ khác "B NOT A".

Ví dụ, khi muốn tìm những tài liệu về chọn giống, nhưng không cần tài liệu về chọn giống ngô, người ta có thể viết biểu thức tìm tin có dạng:

chọn giống NOT ngô

Những biểu ghi được định chỉ số là "**chọn giống**" nhưng không được định chỉ số là "**ngô**" sẽ là biểu ghi phù hợp. Như vậy, tập hợp biểu ghi (tài liệu) thoả mãn yêu cầu tìm chính là hiệu của hai tập hợp: tập hợp biểu ghi thoả mãn yêu cầu có thuật ngữ "**chọn giống**" và tập hợp thoả mãn yêu cầu có thuật ngữ "**ngô**".

Tác dụng của toán tử NOT với biểu thức trên có thể được trình bày bằng sơ đồ Venn như trong hình 7.



Hình 7. Sơ đồ Venn về tác động của toán tử NOT

Trong hình 7, vòng tròn bên trái là tập hợp biểu ghi được định chỉ số với thuật ngữ "**chọn giống**", vòng tròn bên phải là tập hợp biểu ghi được định chỉ số với thuật ngữ "**ngô**". Toàn bộ phần tô đậm là tập

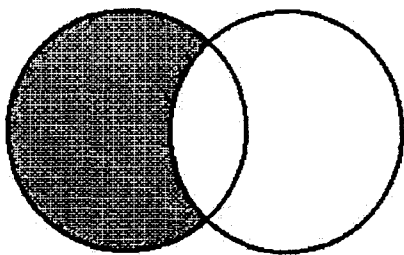
hợp các biểu ghi thoả mãn yêu cầu trong biểu thức tìm. Đó là những biểu ghi được định chi số với thuật ngữ "**chọn giống**" nhưng không được định chi số với thuật ngữ "**ngô**".

Bằng việc phối hợp các toán tử trên đây, ta có thể diễn đạt được một yêu cầu thông tin bất kỳ dưới dạng một biểu thức tìm. Ví dụ, nếu chúng ta có một yêu cầu thông tin là: "*anh/chị hãy tìm những tài liệu về chọn giống ngô cốc, nhưng trừ lúa mì*", chúng ta có thể lập một biểu thức tìm có nhiều thuật ngữ và toán tử trên cơ sở phân tích yêu cầu đó thành các thuật ngữ rồi sử dụng các toán tử để kết hợp như sau:

chọn giống AND (ngô cốc OR lúa OR ngô OR kê) NOT lúa mì

Một đặc điểm cần chú ý là trong tất cả các hệ thống tìm tin tự động hoá, toán tử boole có tác động trên toàn bộ các trường của biểu ghi. Điều đó có nghĩa là những biểu ghi đáp ứng điều kiện của biểu thức tìm, những thuật ngữ có thể nằm trong các trường khác nhau của biểu ghi.

Chẳng hạn ta có biểu thức tìm:



lúa AND chọn giống

Thì nếu biểu ghi có chứa đồng thời hai thuật ngữ trên, nhưng thuật ngữ "*lúa*" có thể nằm trong nhan đề, còn thuật ngữ "*chọn giống*"

có thể nằm trong trường từ khoá hoặc từ chuẩn, thì biểu ghi vẫn được coi là thoả mãn điều kiện tìm và vẫn được tìm ra.

Để trình bày về lý thuyết, người ta thường dùng lý hiệu AND, OR và NOT để chỉ ba loại toán tử boole. Tuy nhiên, ký hiệu cho các toán tử boole trong các hệ thống tìm tin khác nhau có thể khác nhau. Ví dụ về ký hiệu toán tử boole của một số chương trình được trình bày trong bảng 5.

Bảng 5. Ký hiệu toán tử boole của một số chương trình tìm tin

Loại toán tử	Chương trình/Hệ thống	
	CDS/ISIS	DIALOG và STN Messenger
AND	*	AND
OR	+	OR
NOT	^	NOT

Sử dụng toán tử boole trong xây dựng biểu thức tìm tin là phương pháp cơ bản nhất và được áp dụng thường xuyên nhất. Tuy nhiên, để tạo ra những biểu thức tìm đáp ứng được các yêu cầu tìm tin phức tạp còn cần thiết phải sử dụng cả những loại toán tử khác như toán tử lân cận, so sánh,...

2.4. Toán tử lân cận

Hầu hết các hệ thống tìm tin tự động hoá còn cho phép sử dụng một số toán tử khác gọi là toán tử lân cận (proximity operators). Trong một số tài liệu, toán tử lân cận còn có thể được gọi là toán tử

xấp xỉ. Toán tử lân cận là loại toán tử được sử dụng để thể hiện mức độ gần nhau (lân cận) của hai thuật ngữ được viết hai bên toán tử ở trong biểu ghi. Trong khi với toán tử boole, hai thuật ngữ có thể được nằm trong những vị trí rất xa nhau trong cùng một biểu ghi thì toán tử lân cận yêu cầu hai thuật ngữ cùng phải nằm ở những trường như nhau hoặc yêu cầu số từ nằm giữa hai thuật ngữ kẹp hai bên toán tử bằng hoặc không vượt quá một số lượng nhất định.

Chẳng hạn, nếu chúng ta muốn tìm tài liệu về sử dụng năng lượng gió để sản xuất điện và xây dựng biểu thức tìm sử dụng toán tử boole như sau:

năng lượng AND gió

Thì kết quả tìm kiếm có thể đưa ra nhiều tài liệu không phù hợp yêu cầu vì toán tử AND có tác động trong toàn bộ biểu ghi, không đòi hỏi xác định vị trí tương đối giữa hai thuật ngữ nên có thể chúng ta sẽ tìm ra những biểu ghi có chứa cả các thuật ngữ "năng lượng" và "gió" nhưng một thuật ngữ có thể nằm trong trường từ khoá, thuật ngữ kia nằm trong vùng nhan đề. Vì thế sẽ có những biểu ghi tìm ra không phù hợp với yêu cầu tin (với câu hỏi trên, sẽ có thể tìm ra cả những biểu ghi nói về ảnh hưởng của gió đến an toàn đường dây điện,...). Để giải quyết tình trạng đó, chúng ta có thể sử dụng toán tử lân cận để yêu cầu hai thuật ngữ "năng lượng" và "gió" phải đứng gần nhau để tạo ra cụm từ "năng lượng gió" như vậy, chúng ta sẽ có kết quả tìm tin chính xác hơn.

Toán tử lân cận có thể được xếp hai nhóm như sau:

- Toán tử mức câu.
- Toán tử mức trường.

2.4.1. Toán tử lân cận mức câu

Toán tử lân cận mức câu bắt buộc tập hợp biểu ghi thoả mãn yêu cầu phải là những biểu ghi có chứa hai thuật ngữ (tương tự như toán tử AND), nhưng hai thuật ngữ nằm hai bên của toán tử lân cận trong biểu thức tìm phải ở gần nhau (sát nhau) hoặc cách nhau một số từ nào đó trong một câu. Nói cách khác là nó quy định rằng giữa hai thuật ngữ ở hai bên toán tử lân cận có không nhiều hơn một số từ hoặc có chính xác một số từ nhất định.

Ví dụ với chương trình CDS/ISIS, toán tử lân cận mức câu là toán tử . (dấu chấm) và \$ (dấu đôla). Toán tử . (dấu chấm) yêu cầu hai thuật ngữ nằm hai bên của toán tử này phải ở vị trí xa nhau đúng hoặc không quá $(n-1)$ từ, trong đó n là số toán tử dấu chấm được đặt giữa hai thuật ngữ. Ví dụ chúng ta dùng toán tử lân cận . (dấu chấm) của CDS/ISIS để tạo biểu thức tìm như sau:

chọn . giống

thì trên thực tế ta yêu cầu chương trình CDS/ISIS tìm ra các biểu ghi có chứa cụm từ "chọn giống" vì giữa chọn và giống chỉ được cách nhau không quá 0 từ.

Nếu ta có biểu thức tìm có 3 dấu chấm:

hệ...tin

thì những biểu ghi có chứa cụm từ "hệ thống thông tin", "hệ thống tìm tin" đều được coi là thoả mãn điều kiện tìm.

Một số hệ thống tìm tin tự động hoá có những toán tử lân cận còn đòi hỏi trình tự hai thuật ngữ trong biểu ghi phải đúng như thứ tự của thuật ngữ nêu trong biểu thức tìm. Chẳng hạn với hệ thống tìm tin trực tuyến DIALOG hoặc các chương trình tìm tin trên CD-ROM

của DIALOG On DISC, toán tử lân cận (N) không quy định trật tự của hai thuật ngữ trong biểu ghi phải đúng như trong biểu thức, trong khi toán tử (W) lại yêu cầu trình tự của hai thuật ngữ trong biểu ghi phải đúng như trong biểu thức. Ví dụ, nếu với chương trình tìm tin của DIALOG, ta thực hiện hai lệnh tìm với hai biểu thức tìm:

information (N) technology (1)

và

information (W) technology (2)

thì kết quả tìm sẽ khác nhau bởi với yêu cầu tìm thứ nhất, những biểu ghi có cụm từ "**information technology**" và "**technology information**" đều được coi là thoả mãn điều kiện tìm, trong khi với câu hỏi thứ hai, chỉ biểu ghi có chứa cụm từ "**information technology**" mới được coi là phù hợp.

2.4.2. Toán tử lân cận mức trường

Loại toán tử lân cận mức trường yêu cầu biểu ghi tìm được phải chứa cả hai thuật ngữ nằm hai bên toán tử, đồng thời hai thuật ngữ đó phải ở cùng trong một trường hoặc một trường con, phụ thuộc vào loại toán tử.

Những chương trình tìm kiếm khác nhau có thể những loại toán tử lân cận khác nhau. Ví dụ về những dạng toán tử lân cận sử dụng trong chương trình CDS/ISIS và hệ thống tìm tin trực tuyến DIALOG và ý nghĩa của chúng được trình bày trong bảng 6 và 7.

Một số hệ thống còn cung cấp cả loại toán tử lân cận dạng loại trừ (NOT PROXIMITY). Toán tử dạng (NOT PROXIMITY) yêu cầu hai thuật ngữ phải không được ở gần nhau hoặc trong cùng một

trường. Hệ thống DIALOG cung cấp toán tử lân cận dạng (NOT W) và (NOT N). Ví dụ biểu thức tìm có thể viết như sau:

ECONOMIC (NOT W) RECOVERY

Bảng 6. Những toán tử lân cận của CDS/ISIS

Toán tử	Ví dụ	Ý nghĩa
.(dấu chấm)	solar ...energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở sát nhau.
... (nhiều dấu chấm)	solar ... energy	Yêu cầu giữa hai thuật ngữ SOLAR và ENERGY có không quá (n-1) từ, trong đó n là số dấu chấm.
\$	solar \$ energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở sát nhau.
nhiều dấu \$	solar \$\$\$ energy	Yêu cầu giữa hai thuật ngữ SOLAR và ENERGY phải có đúng (n-1) từ (không nhiều hơn, không ít hơn), n là số dấu \$.
(F)	solar(F)energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở trong cùng một trường. Nếu có trường con phải trong cùng một trường con. Nếu là trường lập, phải trong cùng một lần lập.
(G)	solar (G)energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở trong cùng một trường. Có thể ở trường con khác nhau hoặc lần lập khác nhau.

Bảng 7. Toán tử lân cận của chương trình DIALOG

Toán tử	Ví dụ	Ý nghĩa
(W)	solar (W)energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở sát nhau trong biểu ghi tìm được. Thuật ngữ SOLAR phải đứng trước ENERGY trong biểu ghi.
(nW)	solar(3W)energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở cách nhau không quá 3 từ. Thuật ngữ SOLAR phải đứng trước ENERGY.
(N)	solar (N) energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải ở sát nhau. Hai thuật ngữ SOLAR và ENERGY có thể ở bất kỳ thứ tự nào trong biểu ghi.
(nN)	solar(3N)energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY ở cách nhau không quá 3 từ. Hai thuật ngữ SOLAR và ENERGY có thể ở bất kỳ thứ tự nào.
(L)	solar (L) energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải cùng trong trường từ chuẩn của CSDL.
(S)	solar (S) energy	Yêu cầu hai thuật ngữ SOLAR và ENERGY phải trong cùng một trường con của một trường của CSDL.

2.5. Sử dụng ký hiệu chặt cụt

Phương pháp tìm tin bằng kỹ thuật chặt cụt (tiếng Anh là Truncation) là một kỹ thuật tìm tin khá phổ biến và có ý nghĩa quan trọng trong tìm tin tự động hoá. Tìm tin dùng ký hiệu chặt cụt dựa trên khả năng của hệ thống tìm tin so sánh phần gốc của thuật ngữ tìm, từ đầu cho đến phần có dấu chặt cụt với những thuật ngữ có trong tệp đảo của cơ sở dữ liệu. Khi thuật ngữ trong biểu thức tìm tin có ký hiệu chặt cụt, chúng ta sẽ có thuật ngữ chặt cụt và hệ thống sẽ thực hiện phép tìm tin theo kỹ thuật chặt cụt. Trong trường hợp đó, hệ thống tìm tin sẽ tìm những biểu ghi của cơ sở dữ liệu chứa những điểm truy cập có phần gốc từ trùng với phần thuật ngữ chặt cụt trong biểu thức tìm.

Ví dụ, nếu lập biểu thức tìm tin với chương trình CDS/ISIS và sử dụng ký hiệu chặt cụt \$ như sau:

lúa \$

Khi đó những biểu ghi có chứa các thuật ngữ tìm tin như :

lúa

lúa chịu hạn

lúa chịu úng

lúa giống

lúa lai

lúa nếp

lúa nương

lúa xuân

đều được coi là thoả mãn yêu cầu tìm và đều được tìm ra.

Ký hiệu chặt cụt có tác dụng mở rộng kết quả tìm, hỗ trợ tìm kiếm theo số ít và số nhiều.

Đa số hệ thống hỗ trợ phép tìm sử dụng kỹ thuật chặt cụt phía phải. Cũng có hệ thống còn hỗ trợ việc sử dụng ký hiệu chặt cụt ở bên trong từ và sử dụng làm ký tự đại diện, thay thế bất kỳ ký tự nào ở vị trí đó của từ. Ví dụ, trong hệ tìm tin trực tuyến DIALOG, người ta sử dụng dấu hỏi (?) làm ký hiệu chặt cụt. Nếu ta có biểu thức tìm:

organi?ation

thì những biểu ghi có chứa những thuật ngữ:

organisation (theo tiếng Anh)

organization (theo tiếng Mỹ)

đều được coi là phù hợp.

2.6. Tìm so sánh

Phép tìm so sánh cho phép tìm theo các dấu hiệu định lượng. Tìm so sánh còn được gọi là tìm theo giá trị số (numeric searching) hoặc theo dãy (range searching). Toán tử này được sử dụng trong nhiều hệ thống tìm tin. Nó được áp dụng để tìm theo giá trị số lượng hoặc dãy số.

Ví dụ, khi sử dụng chương trình tìm tin CDS/ISIS, để tìm so sánh ta có thể tạo biểu thức tìm tin như sau:

?val(v260) >=1990

trong đó:

? thông báo sử dụng lệnh tìm so sánh

val Hàm val, biến chuỗi ký tự thành số

v260

Trường 260 (năm xuất bản)

>=

toán tử so sánh (lớn hơn hoặc bằng)

Bảng... trình bày toán tử so sánh sử dụng trong Hệ thống DIALOG.

Bảng 8. Toán tử tìm so sánh của DIALOG

Ký hiệu	Ý nghĩa	Ví dụ
=	bằng	SA=1,000,000
<	nhỏ hơn	EM<500
<=	nhỏ hơn hoặc bằng	100<=EM<500 100<=EM<500
>	lớn hơn	SA>1,000,000
>=	lớn hơn hoặc bằng	PY>=1999

2.7. Tìm giới hạn theo trường

Tìm giới hạn theo trường là một kỹ thuật có ý nghĩa khi chúng ta muốn đạt được độ chính xác cao hơn hoặc muốn hạn chế số lượng biểu ghi tìm được.

Thông thường nếu không có những chỉ dẫn đặc biệt về tìm giới hạn theo trường thì hệ thống thực hiện tìm trong mọi trường có thể tìm được. Điều đó có nghĩa là thuật ngữ trong biểu thức tìm có thể thuộc bất cứ trường dữ liệu nào của biểu ghi.

Để tìm giới hạn theo trường phải có những chỉ dẫn cụ thể. Những hệ thống khác nhau sẽ có những chỉ dẫn khác nhau. Trong một số hệ thống, người ta sử dụng tiền tố hoặc hậu tố hoặc toán hạng để thông báo tìm tin giới hạn theo trường.

2.7.1. Tiền tố

Tiền tố mà mã được gán trước thuật ngữ tìm tin để thông báo tìm hạn chế theo trường. Chẳng hạn, trong hệ DIALOG, sử dụng một số tiền tố như AU=, TI=, PY=,...

Ví dụ, những biểu thức tìm dưới đây có ý nghĩa như sau:

Biểu thức tìm **Ý nghĩa**

AU=Smith? Tìm các tác giả có họ là SMITH

TI=energy tìm các biểu ghi, trong nhan đề có chứa thuật ngữ ENERGY

PY=1999 tìm biểu ghi có năm xuất bản là 1999

7.7.2. Hậu tố

Hậu tố là mã gán sau thuật ngữ tìm tin để thông báo tìm giới hạn theo trường. Ví dụ sử dụng hậu tố trong hệ DIALOG trình bày dưới đây:

Biểu thức tìm **Ý nghĩa**

Smith? INAU Tìm các tác giả có họ là SMITH

energy IN TI tìm các biểu ghi, trong nhan đề có chứa thuật ngữ ENERGY

1999 IN PY tìm biểu ghi có năm xuất bản là 1999

Trong chương trình CDS/ISIS, người ta sử dụng toán hạng có công thức:

/(ID)

làm hậu tố, trong đó ID là dấu nhận biết trường do tệp Bảng chọn trường (FST) quy định. Ví dụ, nếu trường tác giả được xác định trong tệp FST có dấu nhận biết trường (ID) là 210, thì lệnh tìm tác giả

Nguyễn Văn Hiệu, giới hạn trong trường tác giả có thể có công thức như sau:

Nguyễn Văn Hiệu/(210)

2.8. Trình tự xử lý và thay đổi mức ưu tiên

Khi trong một biểu thức tìm tin có nhiều loại toán tử khác nhau thì các hệ thống tìm tin xử lý theo thứ tự mức ưu tiên giữa các toán tử đó.

Quy tắc chung của mức ưu tiên của các toán tử thông thường như sau:

- Toán tử lân cận có mức ưu tiên cao hơn toán tử boole và được xử lý trước. Toán tử lân cận mức câu thường được ưu tiên hơn mức trường.

- Toán tử AND và NOT được ưu tiên hơn toán tử OR.

- Khi có các toán tử ngang nhau về mức ưu tiên, hệ thống sẽ xử lý từ trái sang phải.

Ví dụ, mức độ ưu tiên của các toán tử trong CDS/ISIS và DIALOG như sau:

Mức ưu tiên	Toán tử của CSD/ISIS	Toán tử của DIALOG
Cao nhất	. và \$ (F) và (G) * và ^	(W), (N), (L), (S) NOT AND
Thấp nhất	+	OR

2.9. Sử dụng dấu đóng/mở ngoặc đơn

Dấu đóng mở ngoặc đơn có tác dụng nhóm các thuật ngữ để thay đổi thứ tự ưu tiên thực hiện toán tử trong biểu thức. Việc sử dụng

dấu ngoặc đơn cho phép tạo ra những biểu thức tìm tin phức tạp. Nguyên tắc cơ bản của các hệ thống tìm tin là xử lý những kết hợp có ưu tiên cao nhất trước, ưu tiên thấp sau và trong ngoặc trước, ngoài ngoặc sau (tương tự như phép tính số học theo nguyên tắc nhân, chia trước, cộng trừ sau, trong ngoặc trước, ngoài ngoặc sau). Nếu bên trong ngoặc có các toán tử với mức ưu tiên khác nhau thì sẽ thực hiện theo thứ tự ưu tiên chung. *Để đảm bảo cú pháp của hệ thống, số lượng ngoặc mở phải cân bằng với ngoặc đóng.*

Ví dụ nếu có biểu thức tìm tin cho cơ sở dữ liệu CDS/ISIS như sau:

$$(A + B) * (C + D * E)$$

Trình tự xử lý của chương trình sẽ như sau:

#1 $D * E$

#2 $A + B$

#3 $C +$ kết quả của bước 1

#4 Kết quả bước 2 * kết quả bước 3

Nhiều hệ thống cho phép sử dụng nhiều cấp đóng/mở ngoặc. Biểu thức tìm kiếm đó gọi là biểu thức tìm lồng ghép (nested expression).

3. NGÔN NGỮ LỆNH TRONG TÌM TIN TỰ ĐỘNG HOÁ

Để tìm tin trong những hệ thống tìm tin tự động hoá, người tìm tin phải biết sử dụng các lệnh tìm của hệ thống. Ngôn ngữ lệnh của hệ thống tìm tin là một dạng ngôn ngữ hình thức hoá, có cú pháp và vốn từ vựng riêng, có cấu trúc và ngữ nghĩa riêng. Mỗi hệ thống tìm tin khác nhau có thể có hệ thống lệnh khác nhau. Người ta tổng kết là có một số hệ thống lệnh tìm tin như sau:

- Hệ thống dòng lệnh.
- Hệ thống thực đơn.
- Hệ thống hỗn hợp.
- Hệ thống tựa web.

3.1. Hệ thống dòng lệnh

Hệ thống dòng lệnh là hệ thống lệnh tìm tin phát triển đầu tiên. Hệ thống tìm tin theo dòng lệnh được sử dụng cho các hệ thống tìm tin thế hệ đầu tiên. Mọi giao tiếp giữa người sử dụng và hệ thống được thực hiện thông qua ngôn ngữ lệnh của hệ thống. Ngôn ngữ lệnh của hệ thống tìm tin có vốn từ vựng không nhiều, hạn chế. Các lệnh được thể hiện bằng những từ hoặc chữ cái nhất định.

Ví dụ, hệ thống tìm tin trực tuyến DIALOG sử dụng từ "BEGIN" hoặc chữ B để chọn cơ sở dữ liệu. Danh mục một số lệnh của DIALOG được trình bày trong bảng 8.

Bảng 8. Một số lệnh của hệ thống DIALOG

Lệnh	Thí dụ và ý nghĩa
BEGIN hoặc B	BEGIN 208 B 208 (Chọn cơ sở dữ liệu 208 - mã số này tương ứng với cơ sở dữ liệu (file) số 208 trong danh mục cơ sở dữ liệu của ngân hàng dữ liệu DIALOG).
SELECT	SELECT information technology

hoặc S	S information technology (Tìm tin với biểu thức information technology) SELECT S1 AND energy SS1 AND energy (Tìm tin với biểu thức kết hợp giữa lệnh tìm số 1 và energy)
EXPAND hoặc E	EXPAND energy E energy (Xem từ điển thuật ngữ, bắt đầu với từ energy)
DISPLAY hoặc D	DISPLAYS1 DS1 (Xem kết quả của lệnh tìm số 1)
TYPE hoặc T	TYPE S1 TS1 (Xem kết quả của lệnh tìm số 1)
PRINT hoặc P	PRINT S1 PS1 (In kết quả của lệnh tìm số 1)

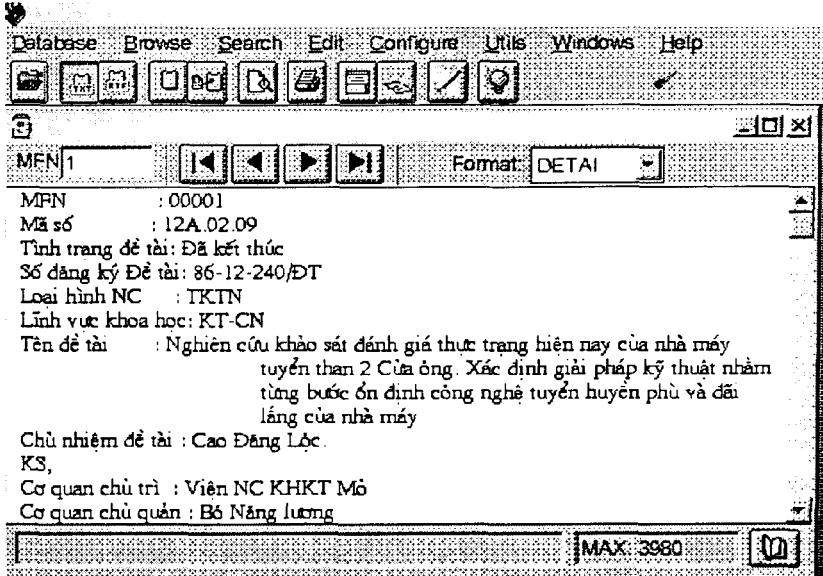
Dòng lệnh có ưu điểm là cung cấp độ linh hoạt cao trong việc xây dựng biểu thức tìm; có khả năng tạo ra những biểu thức tìm linh hoạt, phức tạp, chính xác. Tuy nhiên, nhược điểm cơ bản của dòng lệnh là đòi hỏi người sử dụng phải nhớ (thuộc lòng) các lệnh của hệ thống, phải có nhiều kinh nghiệm và cần được đào tạo cẩn thận.

3.2. Hệ thống thực đơn

Hệ thống thực đơn (Menu) được phát triển sau hệ thống dòng lệnh. Trong hệ thống thực đơn, người sử dụng được yêu cầu lựa chọn các phương án mà hệ thống giới thiệu. Hệ thống thực đơn thường được thiết kế theo cấu trúc thứ bậc (hoặc hình cây). Mức cao nhất khi khởi tạo chương trình tìm được gọi là thực đơn chính (Main Menu). Ví dụ, khi khởi động chương trình khai thác cơ sở dữ liệu của DIALOG Ondisk.exe, chúng ta sẽ thấy thực đơn chính như sau:

Select Search Mode
Easy Menu Search
<i>Dialog Command Search</i>
Online Search
Setup and Accounting
Exit to DOS

Tại thực đơn chính, người sử dụng sẽ được chỉ dẫn đến thực đơn cấp thấp hơn. Để chọn thực đơn tiếp sau, người ta dùng phím con trỏ đưa đến thực đơn cần chọn rồi bấm "ENTER" để chọn. Người ta có thể chọn thực đơn bằng cách bấm các phím đại diện (ví dụ với chương trình CDS/ISIS, chọn phím S tại thực đơn chính để vào menu tìm tin).



Hình 8. Thực đơn chính của chương trình Winisis

Ưu điểm cơ bản của hệ thống thực đơn là đơn giản, dễ sử dụng, không đòi hỏi phải đào tạo lâu. Tuy nhiên, nhược điểm của nó là độ linh hoạt để tìm tin bị giảm hơn so với dòng lệnh; mức lựa chọn bị hạn chế trong các lựa chọn đã cung cấp; có thể bị lẫn thực đơn.

Hệ thống thực đơn được phát triển mạnh trong những năm 90 của thế kỷ trước. Đặc biệt với sự phát triển của các ứng dụng trên nền hệ điều hành Windows, hầu hết các chương trình tìm tin đều được phát triển để chạy trên Windows và hỗ trợ phương thức thực đơn. Thí dụ về thực đơn chính của chương trình Winisis for Windows (thường gọi là Winisis) được trình bày trong hình 8.

3.3. Hệ thống hỗn hợp

Hệ thống hỗn hợp là hệ thống ngôn ngữ lệnh có cả thực đơn và có cả dòng lệnh. Người ta có thể chọn một trong hai phương án.

3.4. Hệ thống tựa Web

Với sự phát triển của Internet và trình duyệt web, ngày càng có nhiều cơ sở dữ liệu được khai thác thông qua World Wide Web. Những hệ thống này được gọi là hệ thống tựa Web (Web-based), sử dụng trình duyệt Web (như Internet Explorer hoặc Netscape Navigator) để làm giao diện tìm kiếm. Với giao diện Web, người tìm tin sử dụng chuột hoặc các thanh cuộn (scrolling bar) để chọn các chức năng theo yêu cầu. Biểu thức tìm tin được nhập vào các biểu mẫu được thiết kế sẵn. Nhiều yếu tố tìm kiếm hoặc các toán tử có thể được lựa chọn bằng các nút chọn hoặc bằng kích chuột vào ô đánh dấu.

Nhiều cơ sở dữ liệu trên CD-ROM cũng được thiết kế để có thể sử dụng trình duyệt Web làm giao diện tìm kiếm trong CSDL. Thí dụ thư viện điện tử IEL của Viện Kỹ sư điện/điện tử quốc tế (IEE/IEEE Electronic Library) sử dụng chương trình tìm tin với sự trợ giúp của trình duyệt Web.

Các dịch vụ thông tin trực tuyến cũng chuyển đổi các chương trình tra cứu cũ sang dạng tựa Web. Ví dụ, Hãng DIALOG, ngoài chương trình tìm tin bằng dòng lệnh thông qua Internet (gọi là **Dialogclassic Web**) còn có chương trình tìm tin tựa Web (Dialog Web). Mạng thông tin quốc tế STN xây dựng chương trình tìm kiếm

gọi là STN Easy với giao diện Web để giúp người dùng tin khai thác ngân hàng dữ liệu của mình (Chương trình tìm tin truyền thống của STN được đặt tên là Messenger và sử dụng dòng lệnh).

Ưu điểm của hệ thống tựa Web là rất dễ sử dụng bởi những chức năng chung như in, lưu kết quả, ... được dựa vào các công cụ của trình duyệt Web (ví dụ như Internet Explorer, Netscape Navigator,...). Những trình duyệt này ngày nay đã trở nên rất quen thuộc với người sử dụng, không mất thời gian tìm hiểu, đào tạo sử dụng.

4. NHỮNG BƯỚC TÌM TIN TỰ ĐỘNG HOÁ

Những kiến thức cơ bản về quá trình tìm tin đã được trình bày ở phần trên. Trong phần này, chúng tôi chỉ trình bày những vấn đề liên quan trực tiếp đến quá trình tìm tin tự động chưa được trình bày chi tiết ở phần trên như phân tích khái niệm, xây dựng biểu thức tìm, chiến lược tìm tin, hiệu chỉnh lệnh tìm,...

Quá trình tìm tin tự động hoá, dù đó là quá trình có hoặc không có chuyên gia tìm tin trung gian có thể được chia thành nhiều bước theo một trình tự hợp lý. Tìm tin tự động hoá là một quá trình truyền thông giữa người có nhu cầu tìm tin và hệ thống tìm tin. Quá trình tìm tin tự động hoá ở những loại hình tìm tin khác nhau có thể khác nhau. Quá trình tìm tin gián tiếp thông qua chuyên gia tìm tin cũng không giống với quá trình tìm tin của người dùng tin trực tiếp. Quá trình tìm tin trên những hệ thống cục bộ, CD-ROM có thể khác với quá trình tìm tin trực tuyến.

Quá trình tìm tin bao giờ cũng bắt đầu bằng một câu hỏi của người có nhu cầu tin. Trong nhiều trường hợp, người dùng tin có thể không thể hiện đúng nhu cầu tin của mình trong câu hỏi gửi tới người tìm tin. Người chuyên gia tìm tin phải tiếp xúc với người dùng tin để làm rõ nhu cầu thông qua quá trình được gọi là phỏng vấn tham khảo (reference interview). Trong quá trình phỏng vấn tham khảo, phải xác định rõ mục tiêu mà người dùng tin muốn đạt được. Mục tiêu tìm tin của người dùng tin có chia thành ba dạng:

- Đạt độ đầy đủ cao về thông tin tìm kiếm: tìm được càng nhiều càng tốt thông tin về chủ đề quan tâm;
- Đạt độ chính xác cao về thông tin tìm kiếm: tìm được những thông tin trùng hợp nhất với chủ đề quan tâm;
- Đạt được sự chọn lọc cao về thông tin: tìm được những bài theo đúng mẫu đã chọn .

Việc xác định đúng yêu cầu tin, mục tiêu tìm kiếm, sẽ giúp cho việc xác định chiến lược tìm phù hợp, thoả mãn cao nhất được nhu cầu tin của người dùng tin.

Khi đã xác định được yêu cầu cụ thể của người dùng tin, người cán bộ tìm tin phải xác định xem cần tiến hành tìm tin trong hệ thống nào. Ngày nay có rất nhiều hệ thống tìm tin tự động hoá, nhiều cơ sở dữ liệu, dịch vụ trực tuyến, cơ sở dữ liệu trên Internet, ... có thể cung cấp thông tin. Vì vậy cần phải xác định để lựa chọn chính xác nơi cần truy cập và thực hiện lệnh tìm. Quyết định lựa chọn cơ sở dữ liệu có vai trò quan trọng vì nó đảm bảo một phần quan trọng trong việc đáp ứng yêu cầu tin. Việc lựa chọn cơ sở dữ liệu phải dựa vào một số tiêu

chí như: diện bao quát đề tài, số lượng nguồn tin được xử lý (số đầu tên tạp chí), mức độ hồi cố, mức độ chọn lọc, mức độ xử lý, giá cả truy nhập,... Quyết định lựa chọn cơ sở dữ liệu thường là của người chuyên gia tìm tin bởi họ là người có hiểu biết hơn về ngân hàng dữ liệu. Người có nhu cầu tin cũng có vai trò nhất định trong việc lựa chọn cơ sở dữ liệu. Họ có thể gợi ý cơ sở dữ liệu tìm kiếm khi họ đã từng sử dụng cơ sở dữ liệu đó.

4.1. Xác định các khái niệm và thuật ngữ tìm

Yêu cầu tìm tin của người dùng tin được thể hiện bằng ngôn ngữ tự nhiên. Yêu cầu đó phải được chuyên gia tìm tin nghiên cứu kỹ càng, phân tích để từ đó rút ra những khái niệm, đối tượng hoặc chủ đề chính của yêu cầu tin. Công đoạn này được coi là việc dịch yêu cầu tin từ ngôn ngữ tự nhiên (ngôn ngữ của người có yêu cầu tin) sang ngôn ngữ tìm tin của hệ thống. Trong hoạt động tìm tin có sự tham gia của người cán bộ tìm tin chuyên nghiệp, người chuyên gia tìm tin là người phân tích khái niệm, dịch các khái niệm sang ngôn ngữ tìm tin của hệ thống và xác định quan hệ giữa những khái niệm đó. Quan hệ này có thể là quan hệ tương giao (phép Và), quan hệ hợp (phép Hoặc), quan hệ loại trừ (phép Trừ).

Những khái niệm đã được rút ra sau khi phân tích yêu cầu tìm tin sẽ được dịch sang thuật ngữ tìm tin sử dụng các công cụ ngôn ngữ tìm tin như: từ điển từ chuẩn, đề mục chủ đề, bộ từ khoá có kiểm soát, bảng phân loại,... Nhằm khắc phục sự lạc hậu của các ngôn ngữ tìm tin, người ta có thể sử dụng từ khoá tự do để thể hiện các khái niệm.

Ví dụ, chúng ta có yêu cầu tin: "Tôi muốn tìm một số thông tin để nghiên cứu có những giống lúa phù hợp cho vùng cao, ít nước tưới, khô hạn". Việc nghiên cứu, phân tích nhu cầu tin cho phép chúng ta có thể rút ra ba khái niệm (chủ đề của yêu cầu là: "lúa", "khô hạn" và "giống". Với những khái niệm trên, chúng ta có thể lập một bảng gồm 3 khối, mỗi khối là tập hợp những thuật ngữ tương đương thể hiện khái niệm (xem bảng 9).

Bảng 9. Phân tích yêu cầu tin thành khái niệm

Khái niệm 1	Khái niệm 2	Khái niệm 3
LÚA	KHÔ	GIỐNG
lúa, lúa nước, lúa nương, oryza sativa	hạn, chịu hạn, khô, vùng cao	chọn giống, chọn lọc, tạo giống

Chúng ta có thể xác định quan hệ giữa các thuật ngữ theo công thức đơn giản sau:

- Quan hệ giữa các khối khái niệm là quan hệ giao, yêu cầu sử dụng toán tử AND;
- Quan hệ giữa các thuật ngữ trong cùng một cột là quan hệ hợp, yêu cầu sử dụng toán tử OR.

4.2. Lựa chọn chiến lược tìm tin

Chiến lược tìm tin có thể được hiểu là một kế hoạch tìm được vạch ra nhằm đạt được mục tiêu tìm kiếm đã được xác định thông qua phân tích nhu cầu tin và mục tiêu tìm tin.

Người chuyên gia tìm tin cần chuẩn bị phương án thay thế khi lệnh tìm được chuẩn bị không cho kết quả như mong đợi. Họ phải chuẩn bị các thuật ngữ thay thế trong trường hợp thuật ngữ được lựa chọn tỏ ra không phù hợp.

Trước khi tìm tin, người cán bộ tìm tin cần xác định rõ các trường có thể giới hạn yêu cầu tìm. Thông thường những hệ thống tìm tin tự động cho phép thực hiện tìm kiếm trong toàn bộ biểu ghi. Tuy nhiên, để nâng cao chất lượng tìm kiếm hoặc tăng độ chính xác hoặc đảm bảo tính đặc thù của mục tiêu tìm tin như xác định trong bước 2, người ta có thể phải thực hiện tìm giới hạn trong một trường nhất định như tìm trong nhan đề, theo tác giả, theo ngôn ngữ, năm xuất bản, từ chuẩn,...

Căn cứ mục tiêu tìm kiếm, người ta đề xuất những chiến lược tìm tin tự động hoá khác nhau. Một số chiến lược cơ bản trong tìm tin tự động hoá bao gồm:

- Chiến lược tìm ngắn gọn;
- Chiến lược khởi xây dựng;
- Chiến lược thu hẹp dần dần;
- Chiến lược mở rộng dần dần.

4.3. Xây dựng biểu thức tìm

Xây dựng biểu thức tìm là một bước hết sức quan trọng. Sau khi đã chuẩn bị kỹ càng các bước trên, người chuyên gia tìm tin sẽ xây dựng một yêu cầu tìm hình thức hoá bằng những thuật ngữ tìm được lựa chọn, sử dụng các toán tử thể hiện quan hệ giữa các thuật ngữ đó.

Trong quá trình tìm tin, có thể xảy ra nhiều tình huống bất ngờ như câu hỏi đặt ra đưa ra quá nhiều kết quả hoặc đưa ra quá ít kết quả, thậm chí không có biểu ghi như mong muốn. Vì thế để đảm bảo cuộc tìm có khả năng thành công cao, người chuyên gia tìm tin cần chuẩn bị sẵn một vài phương án dự phòng như chuẩn bị sẵn các thuật ngữ khác nhau, chuẩn bị phương án thu hẹp khi kết quả tìm ra quá nhiều, v.v...

Sau khi đã chuẩn bị kỹ càng các phương án, người tìm tin truy cập hệ thống và thực hiện lệnh tìm tin theo biểu thức tìm và chiến lược tìm đã vạch ra. Người cán bộ tìm tin cần nắm vững các lệnh của hệ thống, nhập chính xác các lệnh tìm đã chuẩn bị vào hệ thống.

4.4. Đánh giá sơ bộ và hiệu chỉnh kết quả tìm kiếm

Khi thực hiện lệnh tìm xong, người cán bộ tìm tin cần xem nhanh một số biểu ghi đầu tiên để đánh giá nhanh xem lệnh tìm của mình chuẩn bị có đạt yêu cầu không.

Phương pháp đánh giá có thể như sau:

- Đọc nhanh nhan đề.
- Đọc lướt tóm tắt.
- Đọc nhanh trường từ khoá/từ chuẩn.

Trong quá trình đọc đó, người tìm tin cần so sánh kết quả thu được với những kết luận đã rút ra từ bước tìm hiểu nhu cầu tin.

Nếu kết quả tìm kiếm là phù hợp với yêu cầu tin thì tiếp thực hiện các bước tìm kiếm, lưu, tải kết quả về và in,...

Tìm tin tự động hoá là quá trình tìm tin tương tác. Khi đánh giá kết quả trung gian, người cán bộ tìm tin sẽ quyết định tiếp tục thực hiện những bước tiếp theo của lệnh tìm hoặc hiệu chỉnh bằng cách lặp lại những bước trước đó của quá trình tìm tin. Tùy theo tình hình cụ thể, việc xem xét lại có thể từ đầu đến chuẩn bị biểu thức tìm.

Khi xem xét kết quả của lệnh tìm, chuyên gia tìm tin có thể đưa ra những giải pháp để xử lý tình huống nếu lệnh tìm ban đầu không đưa ra kết quả như dự kiến. Có thể có một số tình huống như sau:

4.5.1. Lệnh tìm ban đầu cho ra quá nhiều biểu ghi

Nguyên nhân của tình trạng này có thể khác nhau. Chẳng hạn, do người tìm tin đã sử dụng nhiều từ đồng nghĩa để đảm bảo độ đầy đủ, hoặc sử dụng ít khái niệm để mô tả yêu cầu tin. Để xử lý tình huống này, người tìm tin có thể áp dụng một số phương pháp như sau:

- Đưa thêm những khái niệm mới và dùng toán tử AND để thu hẹp kết quả ban đầu,

- Sử dụng từ chuẩn hoặc từ khoá được chọn ra từ những biểu ghi phù hợp nhất để tăng độ chính xác,

- Nếu dùng nhiều từ đồng nghĩa thì loại bỏ bớt một số từ đồng nghĩa không quan trọng trong lệnh tìm tin,

- Áp dụng phương pháp tìm hạn chế trong một số trường (Ví dụ trong trường từ khoá, trường nhan đề, ...).

4.5.2. Nếu lệnh tìm đưa ra kết quả có quá ít biểu ghi

Nguyên nhân có thể do ta dùng các từ khoá quá chuyên sâu, không thông dụng hoặc sử dụng nhiều khái niệm nhằm đạt kết quả có độ chính xác cao. Để tăng số biểu ghi tìm được, có thể áp dụng những giải pháp sau:

- Loại bỏ một số khái niệm không quan trọng;
- Dùng các thuật ngữ thông dụng hơn thay cho thuật ngữ quá chuyên sâu hoặc không thông dụng;
- Dùng kỹ thuật chặt cụt để nâng cao hơn kết quả;
- Dùng chỉ số phân loại ở cấp cao hơn để làm lệnh tìm.

5. CHIẾN LƯỢC TÌM TIN TỰ ĐỘNG HOÁ

Tìm tin tự động hoá là một hoạt động khoa học, gồm nhiều bước có quan hệ logic với nhau và cần được thực hiện với những chiến lược hợp lý. Trong những tài liệu khác nhau về tìm tin, người ta sử dụng thuật ngữ "chiến lược" theo nhiều cách khác nhau. Có những tài liệu sử dụng thuật ngữ "chiến lược" để thể hiện việc hình thành câu hỏi, biểu thức tìm. Đôi khi người ta sử dụng thuật ngữ "chiến lược" để đề cập đến những hành động, thao tác ở một thời điểm nhất định của quá trình tìm kiếm hoặc đến những quyết định về cách thức nào là tốt nhất để tiếp tục để đạt được mục đích. Trong khuôn khổ chương trình này, chúng ta định nghĩa chiến lược tìm tin là kế hoạch tổng quát được vạch ra hoặc cách tiếp cận nhằm đạt được mục tiêu tìm tin đã được xác định trong quá trình phân tích nhu cầu tin.

Chiến lược tìm tin có thể bao trùm một số bước khác nhau của quá trình tìm tin. Trong xác định chiến lược tìm tin, người tìm tin

phải chú ý và xem xét nhiều vấn đề trước khi tạo lập biểu thức tìm. Những vấn đề này có thể bao gồm:

- Những khái niệm hoặc khía cạnh cần tìm kiếm và thứ tự của chúng;

- Những thuật ngữ phù hợp để thể hiện khái niệm tìm kiếm;

- Những đặc tính của hệ thống tìm tin đang sử dụng;

- Những giải pháp cần tiến hành khi duyệt lại biểu thức tìm.

Để xây dựng được một chiến lược phù hợp, đòi hỏi người tìm tin phải có những hiểu biết về bản chất và tổ chức của cơ sở dữ liệu mục tiêu và nắm bắt được nhu cầu tin của người dùng tin. Sự hiểu biết chính xác về nhu cầu tin cho phép xác định chính xác mục tiêu tìm tin của người dùng tin (Ví dụ như muốn có danh mục đầy đủ tài liệu liên quan, muốn có một số ít tài liệu rất phù hợp với yêu cầu, v.v...). Căn cứ những mục tiêu tìm tin của người dùng tin, có thể có ba kiểu tìm tin:

- **Tìm tin có độ đầy đủ cao:** được tiến hành khi người dùng tin muốn có tất cả những biểu ghi phù hợp về chủ đề quan tâm;

- **Tìm tin với độ chính xác cao:** được tiến hành khi người dùng tin chỉ cần những biểu ghi phù hợp và rất ít nhiều tin (ít biểu không phù hợp);

- **Tìm tin ngắn gọn (brief search):** được tiến hành khi người dùng tin chỉ cần một ít biểu ghi phù hợp mà không cần lấy tất cả biểu ghi phù hợp.

Phụ thuộc vào yêu cầu tin, vào sự hiểu biết của người cán bộ tìm tin về vấn đề tìm tin, về hệ thống tìm tin, về cơ sở dữ liệu được lựa chọn, về phân mềm tìm tin, người chuyên gia tìm tin có thể xác định và chọn một chiến lược phù hợp cho quá trình tìm tin.

Các chuyên gia tìm tin đã đề xuất nhiều chiến lược tìm tin khác nhau. Mỗi chiến lược đều có những đặc thù riêng. Về tổng thể có thể phân loại các chiến lược tìm thành một số dạng chiến lược tổng quát như sau:

- Chiến lược tìm ngắn gọn;
- Chiến lược xây dựng khối;
- Chiến lược các bước liên tiếp (thu hẹp dần dần);
- Chiến lược mở rộng dần dần.

5.1. Chiến lược tìm ngắn gọn

Chiến lược tìm ngắn gọn (Brief search) là cách tiếp cận đơn giản và trực tiếp nhất. Tìm ngắn gọn là cách tìm sử dụng một biểu thức tìm đơn với một số ít thuật ngữ tìm và toán tử boole và có mục đích tìm ra một số biểu ghi phù hợp. Chiến lược này được thường được tiến hành nhằm có được những hiểu biết, ý tưởng cơ bản về cơ sở dữ liệu đang sử dụng để có những điều chỉnh hợp lý sau này.

Chiến lược tìm ngắn gọn thường đưa ra kết quả tìm với độ đầy đủ thấp bởi khái niệm chỉ được trình bày bằng một số ít thuật ngữ.

Ví dụ, với câu hỏi "Tôi muốn tìm một số thông tin để nghiên cứu có những giống lúa tốt cho vùng cao, ít nước tưới, khô han", người tìm tin có thể tạo ra một biểu thức tìm ngắn gọn, đơn giản (với toán tử của CDS/ISIS):

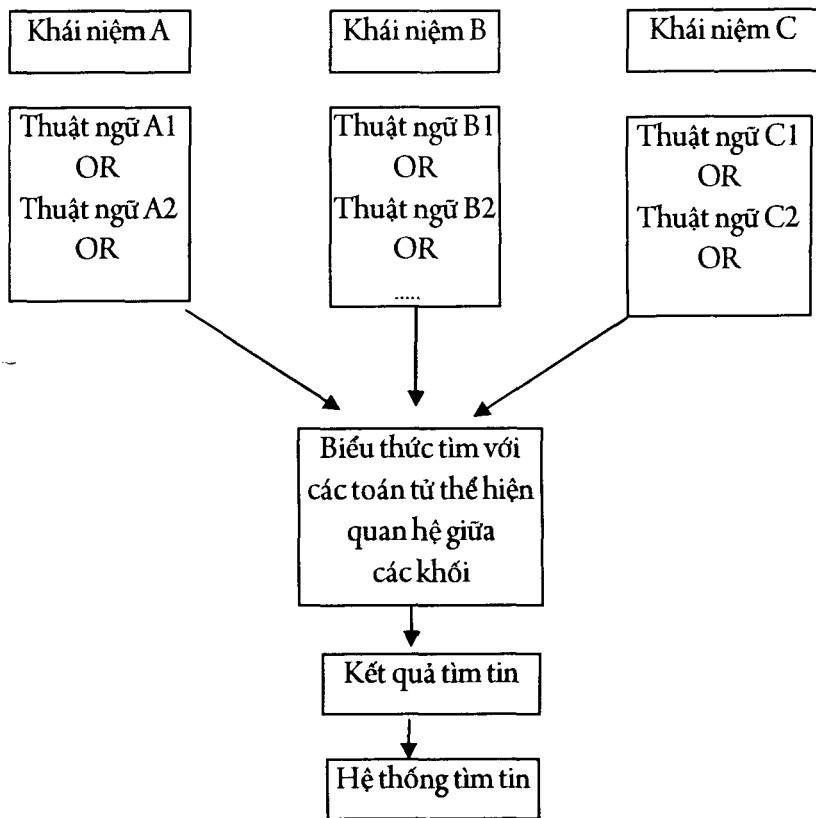
lúa * chọn giống * vùng cao

5.2. Chiến lược xây dựng khối

Chiến lược xây dựng khối (Building blocks strategy) là kế hoạch tìm tin trong đó biểu thức tìm tin được chuẩn bị kỹ và thiết lập chi tiết

ngay từ đầu thông qua việc nghiên cứu kỹ lưỡng yêu cầu tin. Chiến lược này có những đặc trưng sau:

- Mọi khái niệm, thuật ngữ tìm được xác định trước và được thể hiện trong biểu thức ngay từ đầu;
- Những thuật ngữ có quan hệ tương đương được xếp vào khối; phụ thuộc vào yêu cầu tin, có thể có nhiều khối;



Hình 9. Sơ đồ tổng thể chiến lược khối xây dựng

- Các quan hệ giữa các thuật ngữ và giữa các khối được thiết lập từ đầu theo cú pháp của hệ thống tìm tin;
- Biểu thức tìm được thiết kế chi tiết tối đa, viết sẵn thành câu hỏi.
- Mọi bước được chuẩn bị trước từ xem kết quả, lưu kết quả đến in ấn kết quả đều được thiết lập sẵn.

Sơ đồ tổng thể thể hiện chiến lược Khối xây dựng được trình bày trong hình 9.

Với những khái niệm và thuật ngữ được chọn trên, chúng ta có thể tạo ra một biểu thức tìm theo chiến lược Khối xây dựng sử dụng toán tử của CDS/ISIS với các khối như sau:

(lúa + lúa nước + lúa nương + oryza sativa) * (hạn + chịu hạn + khô vùng cao) * (chọn giống + chọn lọc + tạo giống)

Chiến lược xây dựng khối trong thực tế có thể được thực hiện với một lệnh tìm nhưng cũng có thể được thực hiện theo các bước tuần tự: thực hiện tìm lần lượt theo từng khối, sau đó thực hiện lệnh tìm kết hợp kết quả của từng khối. Ví dụ, với câu hỏi đã nêu, chúng ta có thể thực hiện chiến lược tìm theo 4 bước với 4 biểu thức tìm, mỗi biểu thức tương ứng với một khối (một khái niệm):

#1 lúa + lúa nước + lúa nương + oryza sativa

#2 hạn OR chịu hạn OR khô vùng cao

#3 chọn giống + chọn lọc + tạo giống

#4 #1 * #2 * #3

trong đó #1, #2, #3, #4 là ký hiệu câu hỏi thứ 1, 2, 3 và 4 trong chương trình tìm tin CDS/ISIS.

Chiến lược xây dựng khối thực hiện cách tiếp cận tìm tin đảm bảo đồng thời độ đầy đủ và độ chính xác. Chiến lược này được áp

dùng hiệu quả khi người tìm tin hiểu biết hệ thống, làm việc với những yêu cầu tin quen thuộc.

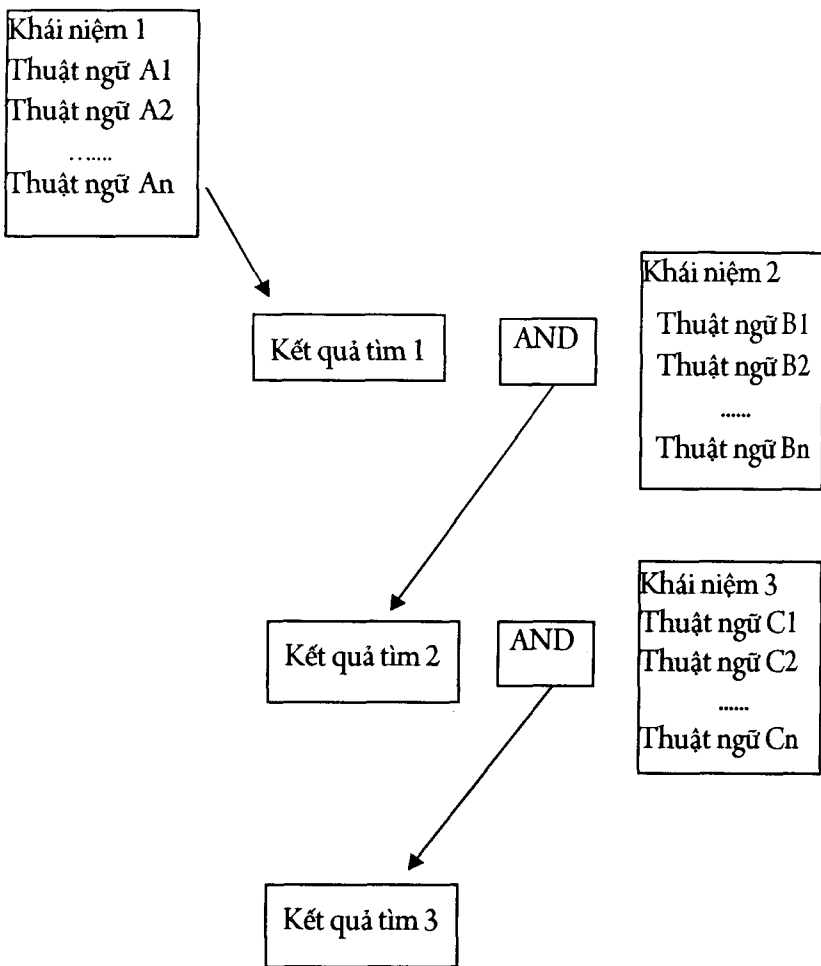
5.3. Chiến lược các bước liên tiếp

Một chiến lược khác cũng thường được áp dụng trong tìm tin tự động hoá là Chiến lược các bước liên tiếp. Chiến lược này còn có thể gọi là chiến lược thu hẹp dần dần. Chiến lược này tìm tin theo tiếp cận thu hẹp dần dần, được đặc trưng với những điểm cơ bản sau:

- Bắt đầu bằng việc tìm theo những khái niệm mức bao quát rộng hơn (với các thuật ngữ tìm có tầm quan trọng nhất trước, các đề mục bao quát rộng hơn) để đảm bảo độ đầy đủ.
- Dựa vào kết quả tìm được thu hẹp dần lại bằng cách chọn từ khoá mới/từ khoá khác, dùng kết hợp AND để thu hẹp dần lại

Chiến lược các bước liên tiếp được gọi là Chiến lược thu hẹp dần dần bởi việc tìm tin sẽ bắt đầu bằng một phương diện tìm kiếm chính, quan trọng nhất, sau đó thu hẹp dần lại bằng biểu thức tìm có kế tiếp kết quả trước sử dụng toán tử AND với khái niệm kế tiếp sau để nâng cao độ chính xác.

Sơ đồ mô tả chiến lược thu hẹp dần được thể hiện trong hình 10. Chiến lược này áp dụng cách tiếp cận đảm bảo độ đầy đủ cao từ đầu, sau đó thu hẹp dần để nâng cao độ chính xác của kết quả.



Hình 10. Chiến lược các bước liên tiếp

5.4. Chiến lược mở rộng dần dần

Một chiến lược tìm tin khác thường được sử dụng khi người dùng tin có nhu cầu về số ít những tài liệu phù hợp, hạn chế số biểu

ghi không phù hợp. Chiến lược mở rộng dần dần từ tài liệu hạt nhân còn được các chuyên gia tìm tin thế giới gọi là Chiến lược "Nuôi ngọc trai" (Growing pearl) (coi tài liệu hạt nhân như một hạt ngọc, từ đó nhân rộng lên). Chiến lược mở rộng từ tài liệu hạt nhân được đặc trưng với một số điểm sau:

- Việc tìm kiếm bắt đầu từ việc tìm một hoặc một vài tài liệu hạt nhân rất phù hợp với yêu cầu tin (có thể là một vài tài liệu đã được gợi ý).

- Dựa vào các tài liệu tìm được trên, hiệu chỉnh biểu thức tìm bằng việc lựa chọn từ khoá hoặc các đặc trưng khác rút ra từ những biểu ghi đã tìm được.

- Kết hợp để nâng dần kết quả tìm tin lên.

Chiến lược này sử dụng cách tiếp cận đảm bảo độ chính xác cao trước tiên, sau đó mở rộng để nâng dần độ đầy đủ.

Trong thực tế, người cán bộ tìm tin phải linh hoạt trong việc áp dụng chiến lược tìm tin.

6. TÌM TIN TRỰC TUYẾN

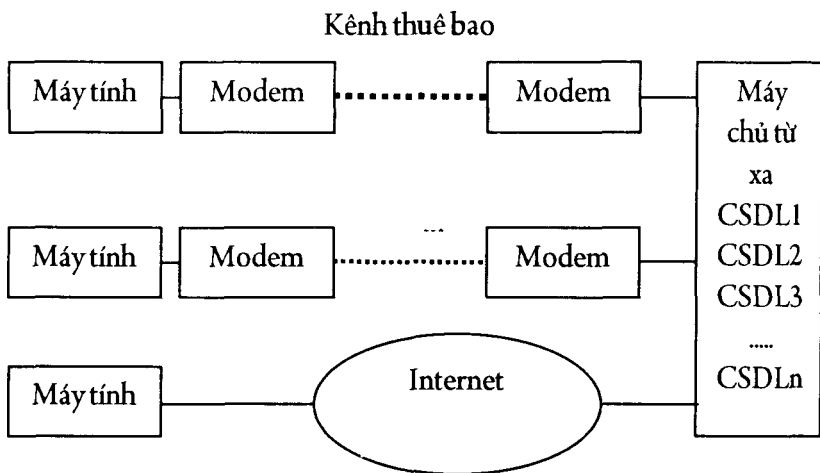
Tìm tin trực tuyến (tiếng Anh là Online Information Retrieval) là quá trình tìm tin tự động hoá mà trong đó người tìm tin sử dụng máy tính điện tử để giao tiếp với dịch vụ thông tin trực tuyến ở xa để thoả mãn nhu cầu tin của mình.

Người tìm tin trực tuyến không nhất thiết phải là người có nhu cầu tin. Họ có thể là người cán bộ tìm tin trung gian thực hiện dịch vụ tra cứu để đáp ứng nhu cầu tin của người dùng tin. Khi người tìm tin là

người trực tiếp có nhu cầu tin, họ là người dùng tin trực tiếp hoặc còn có thể gọi là người dùng tin đầu cuối (end user).

Tìm tin trực tuyến được thực hiện thông qua máy tính điện tử hoặc thiết bị đầu cuối (terminal) kết nối với máy tính chủ trung tâm từ xa thông qua hệ thống mạng máy tính và mạng truyền thông.

Sơ đồ giản lược thể hiện liên kết giữa máy tính điện tử đầu cuối và máy chủ từ xa được thể hiện qua hình 11.



Hình 11. Sơ đồ đơn giản kết nối máy tính và máy chủ từ xa

Máy tính điện tử có thể kết nối với máy chủ từ xa thông qua kênh điện thoại công cộng (phương pháp quay số, dial-up), kênh thuê bao riêng (leased line), qua mạng máy tính hoặc Internet. Vào thời kỳ đầu khi ngành công nghiệp thông tin trực tuyến mới phát triển, các thiết bị đầu cuối không có khả năng xử lý dữ liệu. Việc xử lý dữ liệu dựa hoàn toàn vào năng lực của máy chủ từ xa. Những thiết bị đầu cuối như vậy được gọi là thiết bị đầu cuối câm (dumb terminal). Với sự phát triển

của công nghệ thông tin, máy tính cá nhân được sử dụng làm các thiết bị đầu cuối. Máy tính cá nhân có năng lực xử lý dữ liệu nên được gọi là thiết bị đầu cuối thông tin.

Máy chủ từ xa có thể là của cơ quan, tổ chức đưa lên mạng thông tin, cho phép người dùng tin truy cập và khai thác dữ liệu trong các cơ sở dữ liệu. Tuy nhiên, máy chủ từ xa cũng có thể là của những dịch vụ thông tin trực tuyến, thực hiện dịch vụ trên cơ sở thu phí. Ngày nay có hàng ngàn dịch vụ như vậy trên thế giới. Thí dụ, dịch vụ DIALOG cho phép truy cập trên 400 cơ sở dữ liệu khác nhau, dịch vụ STN cho phép truy cập hơn 200 cơ sở dữ liệu về khoa học và công nghệ.

6.1. Quá trình phát triển của công nghiệp thông tin trực tuyến

Hoạt động thông tin trực tuyến đã trở thành một ngành công nghiệp có sự gia tăng rất nhanh. Ngành công nghiệp thông tin trực tuyến đã phát triển qua một số giai đoạn chính như sau:

Giai đoạn 1. Giai đoạn hình thành và cung cấp dịch vụ không trực tuyến.

Giai đoạn này bắt đầu vào những năm 1960 khi máy tính điện tử được ứng dụng vào hoạt động thư viện-thông tin. Trong thời kỳ này, máy tính điện tử được ứng dụng vào việc tạo ra các cơ sở dữ liệu trên băng từ. Việc phục vụ thông tin được thực hiện theo phương thức không trực tuyến (theo lô - batch mode). Có thể kể đến một số cơ sở dữ liệu lớn được hình thành trong thời kỳ đầu tiên này như Index Medicus (Mỹ) trong lĩnh vực y học, CA (của Chemical Abstracts Services).

Giai đoạn 2. Triển khai hoạt động cung cấp dịch vụ trực tuyến.

Giai đoạn này diễn ra trong khoảng thời gian từ cuối những năm 1960 đến đầu những năm 1980. Trong giai đoạn này, đã xuất hiện các dịch vụ trực tuyến. Các hệ thống sử dụng thiết bị đầu cuối (terminal) để truy cập đến các cơ sở dữ liệu lưu trữ trên các bộ nhớ của máy tính từ xa với sự giúp đỡ của các cán bộ thông tin chuyên nghiệp. Nhờ các tiến bộ của kỹ thuật, đặc biệt của công nghệ thông tin và truyền thông, việc kết nối được thực hiện qua kênh điện thoại mà không bị sai lệch dữ liệu. Vào thời kỳ này đã xuất hiện các máy tính có bộ xử lý trung tâm mạnh, tốc độ xử lý nhanh, làm việc theo chế độ phân thời (time sharing), nhờ đó cho phép nhiều người đồng thời khai thác cơ sở dữ liệu.

Giai đoạn 3. Giai đoạn phát triển bùng nổ.

Giai đoạn này bắt đầu từ đầu những năm 1980 và đặc trưng với sự phát triển bùng nổ của của các dịch thông tin trực tuyến (Xem bảng 10).

Bảng 10. Sự phát triển ngân hàng dữ liệu

Năm	Số CSDL	Số nhà sản xuất CSDL	Số lượng dịch vụ trực tuyến	Số cổng nối (gateways)
1979/1980	400	221	59	
1980/1981	600	340	93	
1986	2901	1379	454	35
1991	5026	2158	731	106
2001	6657	2386	1600	

Trong giai đoạn này, các dịch vụ thông tin trực tuyến tăng liên tục; xuất hiện các dịch vụ với những phần mềm mạnh, tính năng cao, phức tạp. Sự phát triển của các dịch vụ trực tuyến đã tạo ra những ngân hàng dữ liệu rất lớn như DIALOG, STN, DATASTAR,...

Trong giai đoạn này đã xuất hiện công nghệ CD-ROM, cho phép đưa các cơ sở dữ liệu trực tuyến lên CD-ROM để phổ biến cho người sử dụng đầu cuối, giúp cho việc truy cập những cơ sở dữ liệu trở nên dễ dàng hơn, có thể thực hiện theo phương thức tại chỗ.

Giai đoạn 4. Giai đoạn phát triển mạng Internet.

Giai đoạn này đặc trưng với sự phát triển của mạng Internet, làm cho việc truy cập mạng trở nên dễ dàng hơn nhiều; các dịch vụ trực tuyến đã cho phép người sử dụng truy cập dịch vụ của họ qua mạng Internet; nguồn tin điện tử đã phát triển nhanh; hình thành các tổ hợp thông tin điện tử (Aggregators) và cung cấp khả năng truy cập đến các cơ sở dữ liệu toàn văn.

Ngày nay với sự phát triển của thư viện điện tử, cổng giao tiếp điện tử, các dịch vụ thông tin trực tuyến cung cấp ngày càng nhiều hơn các dịch vụ truy cập cơ sở dữ liệu toàn văn.

6.2. Các thành phần của công nghiệp thông tin trực tuyến

Những thành phần cơ bản trong dịch vụ tìm tin trực tuyến bao gồm:

- Các nhà sản xuất cơ sở dữ liệu;
- Các nhà cung cấp dịch vụ trực tuyến;
- Người dùng tin.

6.2.1. Các nhà sản xuất cơ sở dữ liệu

Các nhà sản xuất cơ sở dữ liệu thông thường là các hội nghề nghiệp lớn (như Hội Hoá học Mỹ), các tổ chức nghiên cứu khoa học lớn (như Trung tâm Nghiên cứu khoa học Quốc gia Pháp,...), các thư viện lớn (như thư viện Quốc hội Mỹ, Thư viện Y học Mỹ, ...). Các tổ chức này từ nhiều năm trước đây đã lãnh trách nhiệm tổ chức nguồn tin để cung cấp cho người dùng tin của mình theo một số lĩnh vực khoa học và công nghệ nhất định. Các nhà sản xuất cơ sở dữ liệu còn là các cơ quan thông tin quy mô quốc gia và quốc tế.

Nhiều cơ quan cung cấp dịch vụ trực tuyến cũng tiến hành xây dựng các cơ sở dữ liệu để cung cấp dịch vụ cho khách hàng.

6.2.2. Các nhà cung cấp dịch vụ trực tuyến

Các nhà cung cấp dịch vụ trực tuyến, đôi khi được gọi là các nhà bán hàng trực tuyến (Vendor), là những tổ chức, doanh nghiệp có năng lực thiết bị tin học, có phần mềm phù hợp cho phép quản trị các cơ sở dữ liệu, cho phép quản lý người dùng tin truy cập khai thác cơ sở dữ liệu. Các nhà sản xuất cơ sở dữ liệu cấp phép cho các nhà cung cấp dịch vụ trực tuyến sử dụng cơ sở dữ liệu hoặc thuê cơ sở dữ liệu để thực hiện các dịch vụ thông tin trực tuyến.

6.2.3. Người dùng tin

Người dùng tin có thể bao gồm hai loại chính:

- Hộ dùng tin tập thể: đó có thể là các trung tâm thông tin, thư viện, tổ chức khoa học và công nghệ, doanh nghiệp,... Hộ khai thác các dịch vụ trực tuyến phục vụ cho nhu cầu thông tin của tổ chức mình hoặc cho khách hàng của mình.

- Người dùng tin trực tiếp: đó là các cá nhân trực tiếp khai thác dịch vụ trực tuyến để thoả mãn nhu cầu tin của chính bản thân mình.

Ngày nay với sự xuất hiện của các cơ sở dữ liệu trên CD-ROM, phát triển của mạng Internet, các phần mềm dễ sử dụng, các nhà cung cấp dịch vụ trực tuyến đang đẩy mạnh cung cấp dịch vụ đến người dùng tin trực tiếp.

Tuy nhiên, người cán bộ tìm tin chuyên nghiệp vẫn đóng vai trò quan trọng trong quá trình tìm tin trực tuyến. Họ phải là những người đào tạo, hướng dẫn, định hướng người dùng tin, là người tư vấn về nguồn tin cho người dùng tin trực tiếp.

Trong hoạt động thông tin trực tuyến còn hai thành phần khác cũng có vai trò quan trọng là các nhà cung cấp dịch vụ truyền thông và các nhà cung cấp phần cứng và phần mềm.

7. TÌM TIN TRÊN CD-ROM

7.1. CD-ROM và cơ sở dữ liệu trên CD-ROM

CD-ROM là một dạng bộ nhớ chỉ đọc được ứng dụng rộng rãi trong hoạt động lưu trữ và tìm tin. Thuật ngữ CD-ROM được tạo ra từ các chữ cái đầu của cụm từ tiếng Anh **C**ompact **D**isc **R**ead **O**nly **M**emory (Bộ nhớ chỉ đọc trên đĩa nén). Những đĩa này được gọi là đĩa nén (compact disc) bởi đĩa CD-ROM thông dụng chỉ có đường kính nhỏ (12cm) nhưng dung lượng nhớ của đĩa lên đến 650Mb (vào thời điểm mới xuất hiện công nghệ đĩa CD-ROM, máy tính cá nhân thông thường có bộ nhớ đĩa cứng khoảng 40-80Mb). Với dung lượng như vậy, một đĩa CD-ROM có khả năng chứa dữ liệu tương đương với 464 đĩa mềm 1,2Mb. Dung

lượng nhớ đó có thể lưu trữ được số ký tự ghi trên 200.000 đến 300.000 trang giấy khổ A4, hoặc tương đương 1 tấn giấy.

Nhờ khả năng lưu trữ thông tin có dung lượng lớn, CD-ROM được sử dụng như các bộ nhớ lưu trữ các loại hình dữ liệu đa dạng từ văn bản, đồ họa, hình ảnh, âm thanh, hình ảnh động, dữ liệu đa phương tiện hoặc bất kỳ một tập hợp nào của các dạng dữ liệu trên.

Đĩa CD-ROM được đọc bằng một thiết bị đọc chuyên dụng gọi là đầu đọc CD-ROM. Có thể có nhiều loại đầu đọc CD-ROM. Đầu đọc CD-ROM đơn lẻ được gắn trực tiếp vào máy tính như một ổ đĩa thông thường. Có những thiết bị đọc CD-ROM được lắp độc lập bên ngoài máy tính và hoạt động như một thiết bị ngoại vi. Nhiều đầu đọc CD-ROM có thể được tích hợp vào một thiết bị chuyên dùng gọi là Tháp CD-ROM (CD-ROM Tower) với nhiều đầu đọc (4 đầu hoặc 7 đầu đọc).

Các thiết bị đọc CD-ROM có thể được tích hợp vào mạng máy tính để có thể khai thác dữ liệu trên đĩa CD-ROM theo chế độ mạng. Người ta cũng đã chế tạo được những thiết bị đặc biệt gọi là Tủ CD-ROM (Jukebox) có khả năng lưu giữ đồng thời hàng trăm đĩa CD-ROM, có thể chọn và đưa đĩa CD-ROM vào khai thác tự động trên mạng theo yêu cầu.

Ngày nay, với sự phát triển của khoa học và công nghệ, bên cạnh CD-ROM, công nghệ đĩa quang còn tạo ra nhiều loại đĩa quang khác như CD-PLUS, CD-R, WORM, DVD và đĩa quang ghi xoá được. CD-PLUS là đĩa quang thường được dùng để lưu nhạc trên những thiết bị nghe nhìn bình thường hoặc được dùng để lưu nhạc và hình ảnh khi nghe bằng nhạc bằng máy tính. CD-R cho phép người dùng

có thể ghi dữ liệu lên đó bằng thiết bị chuyên dụng gọi là CD-Writer. WORM là loại đĩa CD-ROM ghi một lần và đọc được nhiều lần. Thuật ngữ WORM được tạo từ cụm từ tiếng Anh **W**rite **O**nce **R**ead **M**any (ghi một lần đọc nhiều lần). Đĩa WORM có thể được sử dụng để sao lưu dữ liệu rất thuận tiện. Đĩa CD ghi xoá được thực chất là loại đĩa quang-từ tính (Magneto-optical CD) có thể ghi và xoá dữ liệu nhiều lần. Đĩa DVD là loại đĩa quang kỹ thuật số (Digital Versatile Disc), tương tự như CD-ROM nhưng có khả năng lưu trữ dữ liệu rất lớn, tới 4,8 Gigabytes trên mỗi mặt đĩa. Đĩa DVD ngày nay đang trở nên rất thông dụng và được sử dụng nhiều trong việc lưu trữ các thư viện điện tử, tạp chí toàn văn.

Khả năng lưu trữ dữ liệu lớn, độ bền cao, dễ bảo quản, vận chuyển và sử dụng đã cung cấp cho đĩa CD-ROM một ưu thế lớn trong việc lưu giữ các cơ sở dữ liệu, kể cả những cơ sở dữ liệu toàn văn hoặc đa phương tiện.

Cơ sở dữ liệu trên CD-ROM ngày nay rất đa dạng. Nó bao quát các loại hình cơ sở dữ liệu như trên các ngân hàng dữ liệu.

Nhiều nhà sản xuất cơ sở dữ liệu đã đưa toàn bộ hoặc từng phần cơ sở dữ liệu lớn lên trên CD-ROM và phát hành trên thị trường. Ví dụ Hiệp hội Hoá học Mỹ đã sản xuất đĩa CD-ROM cơ sở dữ liệu Chemical Abstracts, một trong những CSDL nổi tiếng thế giới. Hoặc Viện Thông tin Khoa học Quốc gia Pháp đã sản xuất đĩa CD-ROM chứa cơ sở dữ liệu PASCAL từ năm 1987 đến nay với khối lượng khoảng hơn 6 triệu biểu ghi.

Nhiều nhà sản xuất cơ sở dữ liệu đã cung cấp các cơ sở dữ liệu toàn văn trên CD-ROM. Ví dụ, Hiệp hội Kỹ sư Điện và Điện tử

(IEEE/IEE) đã cung cấp một Thư viện điện tử trên CD-ROM (IEL) gồm cơ sở dữ liệu tra cứu và tài liệu toàn văn trên CD-ROM. IEL là một thư viện điện tử lưu trữ dữ liệu thông tin có thể truy cập được thuộc lĩnh vực điện - điện tử. IEL chứa hơn 600.000 tài liệu toàn văn được thu thập từ hơn 12.000 ấn phẩm với hơn 2.000.000 trang các tạp chí, tài liệu hội nghị, hội thảo và các tiêu chuẩn thuộc lĩnh vực điện - điện tử xuất bản từ năm 1988 đến nay và hàng tháng được cập nhật khoảng 25.000 trang. Với chương trình tìm kiếm thông tin trên Web IEL 2.0, người sử dụng có thể thực hiện các lệnh tìm kết hợp linh hoạt giữa tên tác giả, chủ đề, tiêu chuẩn và các tham số khác. Trong chế độ xem thông tin, bạn có thể xem từng bài báo, từng tạp chí cũng như in danh sách kết quả tìm, biểu ghi thư mục và đọc file toàn văn trên format PDF trên máy tính.

Nhờ có công nghệ CD-ROM, nhiều cơ sở dữ liệu trước đây chỉ có thể khai thác theo chế độ trực tuyến nay đã có thể khai thác theo chế độ cục bộ. Vì vậy, một số người quan niệm rằng khai thác Cơ sở dữ liệu trên CD-ROM có thể được gọi là khai thác trực tuyến, tuy nhiên điểm khác biệt là cơ sở dữ liệu không còn là từ xa nữa mà có thể được lưu giữ ngay tại chỗ, không phải sử dụng tới các hệ thống viễn thông đắt tiền.

Số lượng các cơ sở dữ liệu trên CD-ROM được tăng lên liên tục hàng năm. Người ta ước tính từ năm 1985 đến nay, số lượng các cơ sở dữ liệu được đưa ra khai thác trên CD-ROM năm sau tăng gấp 2 lần năm trước.

Sự xuất hiện của cơ sở dữ liệu trên CD-ROM và công nghệ sản xuất đĩa CD-ROM đã có tác động nhất định đến thị trường thông tin trực tuyến. Nhiều cơ quan, đặc biệt là ở những trường đại học, đã

chuyển dần từ khai thác thông tin trực tuyến sang khai thác thông tin trên CD-ROM bởi nhiều ưu thế của cơ sở dữ liệu trên CD-ROM (xem bảng so sánh). Người ta đã nhận thấy ở các nước phát triển việc mua cơ sở dữ liệu trên CD-ROM có những đặc điểm khác biệt phụ thuộc vào loại hình cơ quan.

- Các trường đại học tăng cường mua cơ sở dữ liệu trên CD-ROM và giảm số lần tìm tin trực tuyến.

- Các doanh nghiệp ít mua sản phẩm cơ sở dữ liệu trên CD-ROM mà vẫn dựa vào khai thác trực tuyến.

Điều này có thể giải thích được bởi yêu cầu thông tin của doanh nghiệp và của các trường đại học hay viện nghiên cứu có khác nhau. Doanh nghiệp cần những thông tin có tính cập nhật nhanh, thậm chí thông tin cần được cập nhật tức thời (thông tin thời gian thực).

Việc đưa cơ sở dữ liệu lên CD-ROM cũng có ảnh hưởng rõ rệt đến thị trường các sản phẩm tra cứu trên giấy như Chemical Abstracts, Index Medicus,... Các cơ quan thông tin thư viện có xu hướng mua các sản phẩm cơ sở dữ liệu trên CD-ROM thay cho việc mua các ấn phẩm tóm tắt trên giấy.

7.2. So sánh tìm tin trên CD-ROM với tìm tin trực tuyến

Việc khai thác cơ sở dữ liệu trên CD-ROM có những ưu thế nhưng cũng có những hạn chế nhất định so với khai thác trực tuyến.

7.2.1. Ưu điểm của tìm tin trên CD-ROM

- Chi phí viễn thông: Không mất chi phí viễn thông để truy cập cơ sở dữ liệu từ xa, Không mất chi phí cho thời gian truy cập hệ thống từ xa.

- Phân mềm: Có hầu hết thế mạnh của phần mềm tìm tin, phần mềm dễ cài đặt, có hầu hết các phương thức tìm tin cơ bản (theo dòng lệnh, theo thực đơn).

- Truy cập không hạn chế: cung cấp cho người sử dụng khả năng truy cập không hạn chế cho một năm mua cơ sở dữ liệu.

- Khả năng lưu trữ cao, độ bền đáng kể. Chi phí sao lưu không lớn.

- Có khả năng cho phép nhiều người, kể cả người không chuyên nghiệp tìm kiếm trực tiếp.

- Có thể cung cấp đa dạng những loại hình dữ liệu: văn bản, đồ họa, âm thanh, hình ảnh, hình ảnh động hoặc tổ hợp những loại đó.

- Áp lực tâm lý lên người tìm tin thấp hơn khai thác trực tuyến. Có thể lặp lại nhiều lần các bước tìm tin mà không ảnh hưởng đến phí truy nhập.

- Xem dữ liệu, lựa chọn dữ liệu dễ dàng hơn.

7.2.2. Nhược điểm so với tìm tin trực tuyến

- Phải chi cho việc mua quyền truy cập cơ sở dữ liệu trên CD-ROM hàng năm.

- Chi phí cho mua thiết bị nhiều ổ đọc đồng thời, xây dựng mạng khá lớn.

- Tính cập nhật không cao bằng trực tuyến. Khối lượng dữ liệu truy cập được ít hơn.

- Số lượng cơ sở dữ liệu ít hơn so với trực tuyến (cơ sở dữ liệu trực tuyến đa dạng hơn, cho phép truy cập nhiều cơ sở dữ liệu thuộc nhiều lĩnh vực khác nhau).

- Khai thác theo chế độ mạng cũng đòi hỏi kinh phí cao.
- Sự tương hợp giữ các thiết bị và phần mềm.
- Một số yếu tố dữ liệu trên trực tuyến có thể không có khi đưa lên CD-ROM.

Việc lựa chọn giữa khai thác trên CD-ROM hay khai thác trực tuyến phụ thuộc vào nhiều yếu tố khác nhau như:

- Tần suất sử dụng.
- Diện bao quát đề tài.
- Yêu cầu về độ hồi cố.
- Đòi hỏi về tính cập nhật.

Nếu tần suất sử dụng một số cơ sở dữ liệu nào đó mà cao thì nên mua CD-ROM. Trong trường hợp nếu tần suất sử dụng thấp, nhu cầu không cao thì nên khai thác trực tuyến, như vậy sẽ phù hợp hơn về chi phí và lợi ích.

Nếu diện bao quát đề tài rộng thì cần suy nghĩ đến thuê bao trực tuyến bởi nó cung cấp khả năng tiếp cận nhiều cơ sở dữ liệu hơn.

Ngày nay việc sản xuất cơ sở dữ liệu trên CD-ROM đã trở nên đơn giản hơn nhiều. Do công nghệ sản xuất đĩa CD-ROM ngày càng phát triển nên giá đĩa CD-ROM ngày càng giảm. Giá sản phẩm trên CD-ROM về thực chất là giá thuê bao cơ sở dữ liệu trên đó.

Việc phát triển công nghệ đĩa sản xuất CD-ROM ghi xóa được và đầu đọc/ghi CD-ROM đã giúp cho nhiều cơ quan thông tin thư viện có thể tự sản xuất ra những đĩa CD-ROM cơ sở dữ liệu của mình và phổ biến trên thị trường.

Chương V

Tìm tin trên mạng INTERNET

1. NHỮNG KHÁI NIỆM CƠ BẢN VỀ INTERNET

Trong hoạt động thông tin tin ngày nay, không thể không nói đến Internet, một hệ thống thông tin khổng lồ cung cấp đa dạng những thông tin về mọi lĩnh vực khoa học, kinh doanh, kinh tế, xã hội, văn hoá, nghệ thuật,... Internet là công cụ giúp người ta truy cập được đến các thư viện ảo, mục lục thư viện trực tuyến (OPAC), cơ sở dữ liệu trực tuyến, tạp chí điện tử.

1.1. Định nghĩa Internet

Internet là một mạng thông tin toàn cầu và là một nguồn thông tin khổng lồ của nhân loại. Sự hình thành và phát triển có tính chất bùng nổ của Internet được coi là một trong những thành tựu khoa học và công nghệ vĩ đại của thế kỷ XX. Internet chứa một khối lượng thông tin thuộc mọi lĩnh vực từ tinh tế xã hội, khoa học công nghệ, giáo dục và đào tạo, thương mại, thị trường, giá cả, dự báo thời tiết, thông tin giải trí,...

Lúc đầu Internet được định nghĩa là "mạng của tất cả các mạng máy tính sử dụng giao thức IP (giao thức Internet)". Hiện nay điều này không còn hoàn toàn chính xác nữa vì nhiều mạng máy tính với cấu trúc khác (không dùng IP) nhưng nhờ có các cầu nối đa giao thức (Multiprotocol Gateway) nên vẫn có thể nối kết được vào Internet và sử dụng đầy đủ các dịch vụ thông tin trên Internet.

Ngày 24/10/1995 Hội đồng Mạng liên bang Mỹ (The U.S Federal Networking Council) đã nhất trí thông qua định nghĩa:

"Internet là một hệ thống thông tin toàn cầu mà nó :

- *Được kết nối với nhau một cách logic thông qua một hệ thống địa chỉ IP (Internet Protocol) duy nhất toàn cầu hoặc dựa trên những giao thức mở rộng hoặc tương thích của IP.*

- *Có khả năng hỗ trợ chương trình truyền thông sử dụng giao thức TCP/IP (Transmission Control Protocol/ Internet Protocol, còn gọi là giao thức điều khiển truyền dẫn/ giao thức Internet), hoặc các giao thức tương thích hoặc mở rộng.*

- *Cung cấp, sử dụng hoặc làm cho các dịch vụ cấp cao nằm trên tầng truyền thông hoặc các hạ tầng liên quan có thể truy cập được một cách công cộng hoặc cá nhân".*

Nghị định 55/2001/CP-NĐ của Chính phủ ngày 23/8/2001 về hoạt động Internet đã định nghĩa "Internet là một hệ thống thông tin được kết nối với nhau bởi giao thức truyền thông Internet (IP) và sử dụng một hệ thống địa chỉ thống nhất trên phạm vi toàn cầu để cung cấp các dịch vụ và ứng dụng khác nhau cho người sử dụng".

Internet sử dụng bộ giao thức truyền thông TCP/IP (từ tiếng Anh là Transmission Control Protocol/Internet Protocol) làm chuẩn truyền thông của toàn mạng.

Khi một mạng cục bộ (mạng LAN) sử dụng công nghệ của Internet trong hoạt động của mạng thì được gọi là mạng Intranet.

1.2. Địa chỉ IP và địa chỉ tên miền

Trên Internet có hàng triệu máy chủ được kết nối trên đó có vô vàn các tài nguyên thông tin được phân bố một cách phân tán. Để thông tin có thể được chuyển tải từ một máy này đến một máy khác một cách chính xác, mỗi máy tính kết nối vào mạng đều có địa chỉ và đó là địa chỉ duy nhất, không trùng lặp trên quy mô toàn cầu. Để đảm bảo sự thống nhất và không trùng lặp đó, người ta đã thống nhất những quy tắc về xác định địa chỉ cho máy tính trên mạng.

Địa chỉ trên mạng Internet có hai đặc trưng cơ bản là:

- Tính đẳng cấp.
- Tính duy nhất.

Internet sử dụng hai loại địa chỉ có giá trị ngang nhau để xác định một máy chủ kết nối vào mạng: Địa chỉ IP và Địa chỉ tên miền.

1.2.1. Địa chỉ IP

Mỗi máy tính nối vào mạng Internet đều có một địa chỉ IP riêng và còn được gọi là địa chỉ Internet. Địa chỉ này là duy nhất toàn cầu cho mỗi máy tính nối mạng Internet. Không thể có hai máy tính trên mạng Internet có cùng một địa chỉ IP. Địa chỉ IP là địa chỉ dạng số, gồm có bốn nhóm số cách nhau bằng dấu chấm. Mỗi nhóm số có thể có giá trị từ 0 đến 255.

Ví dụ, địa chỉ máy Web chủ mạng VISTA của Trung tâm Thông tin Khoa học và Công nghệ Quốc gia là:

192.168.2.2

Bốn nhóm số này sẽ thể hiện cả máy chủ cụ thể cả mạng con mà máy chủ đó kết nối vào. Phần nhóm chữ số về mạng con là từ trái sang phải.

Phần nhóm số cho mạng được Trung tâm Thông tin Mạng Internet Quốc tế (InterNIC), theo uỷ quyền của Tổ chức phân phối địa chỉ Internet - IANA (Internet Assigned Numbers Authority), quyết định và gán cho nhà cung cấp dịch vụ Internet. Các nhà cung cấp dịch vụ xác định và gán phần nhóm số cho các máy trong mạng của mình.

Lúc đầu các máy tính nối trên Internet đều sử dụng địa chỉ IP. Tuy nhiên, việc định danh các phần tử của liên mạng bằng các con số như trong địa chỉ IP là không thuận tiện cho người sử dụng bởi chúng rất khó nhớ và dễ gây nhầm lẫn.

Để khắc phục tình trạng này, người ta xây dựng một hệ thống địa chỉ tên miền (domain name) bằng chữ để dùng song song với hệ thống địa chỉ IP.

1.2.2. Địa chỉ tên miền

Địa chỉ tên miền (domain name) là một dạng địa chỉ bằng chữ sử dụng để xác định một máy tính kết nối vào Internet. Địa chỉ này có tính đăng cấp, duy nhất và có địa chỉ IP tương ứng. Địa chỉ tên miền bao gồm một số cụm ký tự và phân cách nhau bằng dấu chấm.

Ví dụ về một số địa chỉ tên miền như sau:

www.vnn.vn

www.vista.gov.vn

Thông thường, trong địa chỉ tên miền có thành phần được gọi là "tên miền cấp cao" (top-level domain name). Với các mạng ngoài nước Mỹ, phần tên miền cấp cao gồm hai bộ phận: bộ phận xác định quốc gia và bộ phận xác định lĩnh vực hoạt động hoặc loại hình của tổ chức, cơ quan hoặc mạng. Miền cấp cao nhất là cấp quốc gia. Mỗi quốc gia được gán một mã tên miền riêng gồm hai chữ cái. Ví dụ:

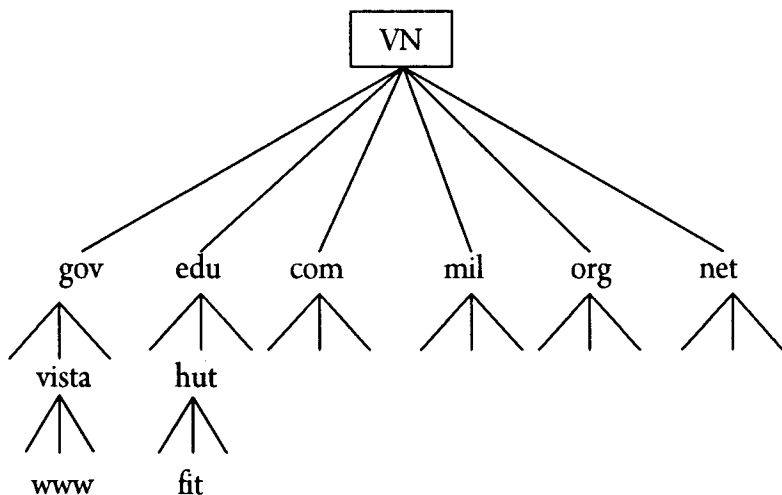
Mã miền cấp quốc gia	Nước
vn	Việt Nam
au	Ôxtralia
ca	Canada
fr	Pháp
jp	Nhật Bản
uk	Anh
ru	Nga
de	Đức
kr	Hàn Quốc
sg	Singapore

Trong từng quốc gia lại được chia thành một số miền cấp cao xác định lĩnh vực hoặc loại hình tổ chức và tiếp tục đi xuống các cấp thấp hơn. Bảng 11 liệt kê một số mã tên miền cấp cao về lĩnh vực.

Bảng 11: Một số miền cấp cao về lĩnh vực

Mã tên miền	Dạng tổ chức
Com	Thương mại, doanh nghiệp, kinh doanh (Commercial)
Edu	Giáo dục
Gov	Nhà nước
Mil	Quân đội
Org	Các tổ chức khác
Net	Tài nguyên trên mạng
Int	Tổ chức quốc tế

Ví dụ, sự phân cấp tên miền điển hình ở Việt Nam có thể được thể hiện ở hình 12.



Hình 12. Sơ đồ phân cấp đơn giản tên miền điển hình ở Việt Nam

Mỗi địa chỉ tên miền bao giờ cũng sẽ có một địa chỉ IP tương ứng. Ví dụ, địa chỉ tên miền của máy chủ Web của VISTA là:

www.vista.gov.vn

Còn địa chỉ IP tương ứng của máy chủ Web của mạng VISTA là:

192.168.2.2

Các máy tính và thiết bị trên Internet sử dụng địa chỉ IP để truyền dữ liệu. Vì thế trên Internet có một hệ thống dịch vụ gọi là Hệ thống tên miền DNS (Domain Name System) có chức năng dịch tên miền sang địa chỉ IP tương ứng để các trạm làm việc với máy chủ bằng cách dùng địa chỉ IP chứ không dùng tên nữa.

1.3. Các dịch vụ cơ bản của Internet

Internet là một kho thông tin phân tán khổng lồ và để khai thác nó, người ta tạo ra công cụ khác nhau. Những công cụ khai thác thông tin trên Internet bao gồm:

- TELNET (Đăng nhập từ xa).
- FTP (Dịch vụ truyền tệp).
- Thư điện tử.
- ARCHIE (Tìm kiếm tệp).
- GOPHER (Tra cứu theo thực đơn).
- WAIS - (Tìm kiếm thông tin theo chỉ số).
- World Wide Web (WWW hoặc Web).

Việc phân loại các dịch vụ trên Internet chỉ có tính chất tương đối. Các dịch vụ trên Internet có thể được phân loại thành các nhóm tùy thuộc vào các tiêu chí phân loại khác nhau.

Nếu dựa vào thời gian xuất hiện, có thể chia các dịch vụ Internet thành:

- Dịch vụ Internet thế hệ 1: bao gồm các dịch vụ TELNET, FTP, thư điện tử, nhóm tin. Đây là những dịch vụ xuất hiện đầu tiên, từ khi xây dựng Internet.

- Dịch vụ Internet thế hệ 2: bao gồm các dịch vụ ARCHIE, GOPHER và WAIS. Những dịch vụ này xuất hiện muộn hơn nhưng đang dần dần bị thay thế bằng dịch vụ thế hệ 3.

- Dịch vụ Internet thế hệ 3: dịch vụ World Wide Web, là dịch vụ mới nhất và bao quát nhất. Dịch vụ Web có khả năng bao quát tất cả các dịch vụ trước đó.

Nếu dựa vào chức năng thông tin, có thể chia thành các nhóm:

- Các dịch vụ lấy thông tin: gồm dịch vụ FTP và GOPHER.
- Các dịch vụ tìm kiếm thông tin: gồm các dịch vụ WAIS và ARCHIE.
- Các dịch vụ truyền thông: Dịch vụ thư điện tử (E-mail) và TELNET.
- Các dịch vụ thông tin đa phương tiện: World Wide Web.

World Wide Web, còn thường được viết tắt là Web hay WWW, là một dịch vụ mới nhất và hấp dẫn nhất của Internet. Đó tập hợp khổng lồ dữ liệu trải trên hàng triệu máy tính tham gia vào mạng Internet toàn cầu. Thậm chí nhiều người quan niệm World Wide Web và Internet là một. Thực tế thì World Wide Web chỉ là một dịch vụ của Internet, nhưng lại là phần phát triển nhanh và được sử dụng

nhieu nhất. Đây là một dịch vụ thông tin có khả năng tích hợp trong nó tất cả các dịch vụ khác của Internet như thư điện tử, FTP, GOPHER và WAIS. Vì vậy khi nói đến tìm tin trên Internet, người ta thường tập trung chú ý đến các công cụ tìm tin trên Web.

Những giới thiệu về quá trình hình thành phát triển cũng như các phương pháp tìm kiếm thông tin trên Web được trình bày chi tiết ở phần sau.

2. WORLD WIDE WEB

2.1. Một số khái niệm cơ bản của World Wide Web

World Wide Web (sau đây viết tắt là Web) dịch vụ Internet được xây dựng dựa trên một kỹ thuật trình bày dữ liệu có tên gọi là *siêu văn bản* (hypertext). Siêu văn bản có thể là một từ hoặc một câu xuất hiện trong các tệp điện tử và được liên kết với thông tin trong các tệp khác. Khi chọn siêu văn bản, màn hình sẽ tự động hiển thị các tệp và các thông tin liên kết với từ hoặc câu đó. Các thông tin liên kết đó có thể là văn bản khác, là hình ảnh, âm thanh, hình ảnh động hoặc là tổ hợp của các dạng đó. Người dùng có thể dễ dàng quay lại tệp chứa siêu liên kết gốc bằng cách sử dụng phím Back. Nếu có sẵn một siêu liên kết khác, người dùng có thể thực hiện ngay việc liên kết với các trang mới. Nói cách khác, siêu văn bản biểu hiện qua các từ được chọn trong văn bản có thể được “mở rộng” bất kỳ lúc nào để cung cấp các thông tin đầy đủ hơn về các từ đó. Sự “mở rộng” ở đây được hiểu là chúng có các liên kết “links” tới các tài liệu khác chứa những thông tin bổ sung. Siêu văn bản được phân biệt với văn bản thường bởi màu sắc và kiểu dáng của nó. Phần mềm dùng để nối đến Web cài đặt trên

máy trạm, gọi là bộ duyệt hay trình duyệt Web (Web Browser), sẽ xác định siêu văn bản đó có dạng như thế nào (thường là màu xanh và được gạch dưới, nói chung là một màu và một kiểu dáng độc nhất để dễ phân biệt với văn bản thông thường).

Thuật ngữ và khái niệm siêu văn bản là do nhà tin học Tel Nelson đưa ra lần đầu tiên năm 1965 trong bài viết "Computer dreams" (những giấc mơ máy tính) bàn về ý tưởng tạo ra những máy tính có "trí tuệ nhân tạo", có tư duy như con người; trong đó ông đã đề cập đến khái niệm siêu văn bản và muốn cả thế giới có một hệ thống siêu văn bản. Đến năm 1980, Tim Berners - Lee, một nhà nghiên cứu làm việc tại Viện Nghiên cứu Hạt nhân châu Âu (CERN) - một tổ hợp phòng thí nghiệm vật lý hạt nhân đóng ở Thụy Sĩ đã giúp cho ý tưởng về siêu văn bản đầu trở thành hiện thực. Năm 1990, cùng với đồng nghiệp là Robert Cailliau, Tim Berners-Lee đã bắt tay vào thực hiện dự án thiết kế mạng thông tin toàn cầu dựa trên khái niệm siêu văn bản. Thuật ngữ World Wide Web được dùng để mô tả dự án và được công bố lần đầu tiên vào tháng 8/1991.

2.1.1. Trang Web và HTML

Web thực chất là một tập hợp các công cụ tiện ích và giao diện giúp người sử dụng có thể tự tạo ra các siêu văn bản và cung cấp cho những người dùng khác trên Internet. Trang siêu văn bản như vậy thường được gọi là "**trang Web**". Để soạn thảo các trang siêu liên kết người ta sử dụng công cụ *Ngôn ngữ đánh dấu siêu văn bản* thường viết tắt là HTML (từ tiếng Anh HyperText Markup Language), do đó các trang Web còn được gọi là tài liệu siêu văn bản hay trang HTML. Ngôn ngữ HTML thực chất là một phần rút gọn của ngôn ngữ đánh

dấu văn bản chuẩn tổng quát SGML (Standard General Markup Language). HTML liên kết các kiểu dữ liệu khác nhau trên cùng một trang tài liệu như văn bản, đồ hoạ, hình ảnh động, âm thanh.

HTML là một ngôn ngữ định dạng (hay đánh dấu - markup). Mỗi tệp văn bản có chứa các mã đánh dấu gọi là các “thẻ” (tag) để định dạng tài liệu hoặc tạo các siêu liên kết. Thẻ HTML là một đoạn mã được giới hạn bởi các dấu ngoặc nhọn “<” và “>”. Ví dụ mã đánh dấu bắt đầu trang Web là <HTML> còn mã đánh dấu kết thúc trang Web là </HTML>. Các thẻ định dạng HTML luôn luôn đi theo cặp. Một mã mở đầu đánh dấu đoạn cần định và mã đánh dấu kết thúc phân cần định dạng. Mã kết thúc gần giống với mã mở đầu, chỉ khác ở chỗ có thêm 1 gạch chéo ở phần đầu.

Ví dụ về một trang tài liệu HTML có thể như sau:

<HTML>

<HEAD>

<TITLE>Ví dụ về trang HTML</TITLE>

</HEAD>

<BODY>

<H1>Đây là một trang Siêu văn bản đơn giản</H1>

HTML là một ngôn ngữ định dạng (hay đánh dấu - markup).

Mỗi tệp văn bản có chứa các mã đánh dấu gọi là các thẻ (tag) để định dạng tài liệu hoặc tạo các siêu liên kết.

</BODY>

</HTML>

Trong đó <HTML> thông báo bắt đầu trang HTML, </HTML> đánh dấu kết thúc trang HTML, <BODY> thông báo bắt đầu phần nội dung sẽ hiển thị trên màn hình còn </BODY> đánh dấu kết thúc phần này.

Chương trình đọc trang HTML và hiển thị nó cho người đọc xem gọi là trình duyệt Web (Web Browser). Trình duyệt Web sẽ đọc các thẻ khi thực hiện định dạng các tệp HTML trên màn hình.

2.1.2. Định vị tài nguyên thống nhất - URL

Để thực hiện việc truy nhập, liên kết các tài nguyên khác nhau của mạng Web theo kỹ thuật siêu văn bản, mỗi tài nguyên trên mạng, kể cả trang web, đều phải có địa chỉ trên Internet. Web sử dụng khái niệm **định vị tài nguyên thống nhất** (có thể còn gọi là **Định vị nguồn tin thống nhất**), viết tắt là URL từ thuật ngữ tiếng Anh **Uniform Resource Locator**) để xác định địa chỉ của một tài liệu, một dịch vụ, máy chủ,... trên Web.

Cấu trúc tổng quát của một URL bao gồm các thành phần :

Giao thức:// địa chỉ máy chủ/ đường dẫn/ tên tệp

trong đó:

- **Giao thức** cho biết loại giao thức mà trình duyệt phải sử dụng để lấy ra tệp tin định trước. Những giao thức phổ biến được dùng trong các URL là:

http giao thức truyền tệp siêu văn bản, truy cập tài nguyên trên máy chủ Web.

ftp giao thức truyền tệp, truy cập tài nguyên trên máy chủ truyền tệp (FTP).

gopher	giao thức gopher, truy cập tài nguyên trên máy chủ Gopher.
telnet	giao thức telnet, thực hiện truy cập đến máy chủ từ xa.
file	truyền tệp cục bộ, truy cập đến tệp tin trên máy cục bộ hoặc trên máy chủ truyền tệp.
wais	giao thức wais, truy cập tài nguyên trên máy chủ Wais.
news	truy cập tài nguyên trên máy chủ quản lý nhóm tin (Usenet News).

Mặc dù tất cả các URL đều bắt đầu bằng một giao thức, nhưng cách biểu diễn giao thức đó có thể thay đổi. Tuy nhiên, hầu hết các giao thức đều được biểu diễn bằng dấu hai chấm và hai gạch chéo (://) ngoại trừ giao thức tin tức và địa chỉ thư là dùng dấu hai chấm (**news:**, **mailto:**)

- **Địa chỉ máy chủ:** là thành phần đứng sau giao thức và là địa chỉ của máy chủ lưu trữ tệp tin. Địa chỉ máy chủ có thể là địa chỉ IP hay địa chỉ tên miền.

Ví dụ địa chỉ máy chủ của mạng VISTA có thể là:

192.168.1.2

hay

www.vista.gov.vn

- **Đường dẫn thư mục** là thành phần đứng sau tên địa chỉ máy chủ (directory path) để thông báo đường dẫn đến thư mục lưu tệp tin. Một điểm cần chú ý là dấu “/” được sử dụng để chỉ ra các thư mục chứ không phải dấu “\” trong môi trường Windows hay DOS.

- Tên tệp là thành phần sau cùng của một URL. Khi không có tên tệp thì trình duyệt Web sẽ sử dụng một tên mặc định do máy chủ

Web cụ thể quy định (thông thường tên ngầm định là **index.htm** hoặc **index.html**).

Để truy cập tới một trang Web nào đó ta có thể gõ trực tiếp URL của trang Web đó vào khung cửa sổ địa chỉ của trình duyệt Web trên máy trạm. Nếu trang Web có những siêu liên kết thì đằng sau mỗi siêu liên kết chính là một URL sẵn sàng để được đưa vào sử dụng. Khi chọn một siêu liên kết, thực tế người dùng đã cung cấp một URL tương ứng với liên kết đó cho trình duyệt của mình. Trình duyệt sẽ tiến hành các công đoạn tìm kiếm và hiển thị tệp có URL đã được xác định bởi lệnh siêu liên kết đó.

2.2. Trình duyệt Web

Hoạt động của Web dựa trên mô hình khách/chủ (tiếng Anh gọi là Client/Server). Các trang Web sẽ được cài đặt trên máy chủ Web (Web Server) và có thể truy cập được qua mạng Internet nhờ chương trình máy khách (trình duyệt).

Để truy cập được các trang Web, trên máy trạm, người dùng cần một chương trình đặc biệt gọi là **Trình duyệt Web** (tiếng Anh gọi là **Web Browser**). Trình duyệt Web là một chương trình máy khách thực hiện chức năng giao tiếp giữa người dùng và các tài nguyên từ xa lưu giữ trên các máy chủ Web với hai chức năng cơ bản:

- Liên lạc với máy chủ và gửi yêu cầu tìm kiếm các tệp html đến máy chủ Web ở xa trên mạng Internet dựa vào địa chỉ URL được khai báo;

- Nhận thông tin từ máy chủ Web, biên dịch các thẻ html và hiển thị nội dung các trang tài liệu nhận được.

Trình duyệt Web có thể chia thành hai nhóm:

- Trình duyệt văn bản;
- Trình duyệt đồ họa.

Trình duyệt văn bản không có khả năng hiển thị đồ họa, hình ảnh, hình ảnh động, âm thanh,... Chúng chỉ cho phép người dùng xem được văn bản. Ví dụ về các trình duyệt văn bản là MOSAIC và LYNX. Đây là những loại trình duyệt thế hệ đầu tiên.

Trình duyệt đồ họa thường có thêm các chương trình trợ giúp hoặc có khả năng tích hợp các chương trình này để cho phép người dùng hiển thị được hình ảnh, chạy các chương trình nghe âm thanh, xem hình ảnh động. Ví dụ về những trình duyệt đồ họa là Netscape Navigator của Netscape, Internet Explorer của Microsoft, Web Access của Novell.

2.3. Nguồn tin trên World Wide Web

Nguồn tin trên Web rất phong phú và đa dạng. Chúng ta có thể xếp các loại thông tin trên Web thành các nhóm như sau:

- **Các tài liệu điện tử:** có hàng tỷ trang Web về mọi lĩnh vực khác nhau được đưa lên mạng. Mọi loại hình tài liệu nếu được chuyển đổi sang dạng số (điện tử) hoặc là dạng số ngay từ khi được tạo ra đều có thể lưu trữ trên Web và có thể truy cập được bởi người truy cập Internet. Những tài liệu này có thể là các tệp văn bản ở những khổ mẫu khác nhau (HTML, Word, văn bản thuần túy, PDF, Postscript,...), các tệp ảnh (ảnh GIF, TIFF, Bitmap, PNG,...); tệp âm thanh, tệp đồ họa, tệp hình ảnh động;

- **Phần mềm máy tính:** Trên Web, người sử dụng Internet có thể truy cập và tải về nhiều phần mềm máy tính. Một số phần mềm là

loại không mất tiền (Freeware), một số khác là loại có thể sử dụng cho mục đích đánh giá và dùng trong một thời gian hạn chế (thường gọi là Shareware). Các phần mềm này rất đa dạng, từ phần mềm hệ thống, phần mềm tiện ích đến phần mềm ứng dụng;

- **Tạp chí điện tử:** với sự phát triển của Web, nhiều tổ chức đã xuất bản các tạp chí và đưa lên mạng để mọi người khai thác, sử dụng. Tạp chí điện tử là một nguồn tin vô cùng quan trọng trên Web. Tạp chí điện tử có thể là dạng điện tử của tạp chí được xuất bản trên giấy hoặc trên vật mang tin khác. Trong nhiều trường hợp tạp chí điện tử không có bản tương tự trên giấy mà chỉ có ở dạng điện tử mà thôi. Những tạp chí điện tử không có phiên bản trên giấy còn được gọi theo tiếng Anh là E-zines hoặc E-Journals. Cùng với các tạp chí điện tử, còn có rất nhiều các báo điện tử cung cấp tin tức và nhiều thông tin khác. Về nội dung, các tạp chí điện tử có thể không khác với phiên bản trên giấy, tuy nhiên điểm vượt trội của tạp chí điện tử là nó có các siêu liên kết cho phép kết nối tài liệu hiện tại với các tài liệu khác. Ngoài ra, tạp chí điện tử có ưu điểm rất lớn là dễ phổ biến và bảo quản dễ dàng;

- **Dịch vụ thông tin trực tuyến:** Ngày nay tất cả các nhà cung cấp dịch vụ thông tin trực tuyến thương mại đều đã cung cấp dịch vụ của mình thông qua Web. Ngoài ra còn rất nhiều dịch vụ trực tuyến miễn phí được các tổ chức xây dựng và cung cấp cho người sử dụng;

- **Các cơ sở dữ liệu:** Nhiều tổ chức đã xây dựng các cơ sở dữ liệu và đưa chúng lên mạng để người dùng tin khai thác. Một số cơ sở dữ liệu có thể là miễn phí trong khi một số cơ sở dữ liệu thì chỉ hạn chế cho những người sử dụng có đăng ký. Những cơ sở dữ liệu này rất đa dạng, có thể là thư mục, dữ kiện hoặc toàn văn;

- **Các mục lục thư viện trực tuyến:** Ngày nay hầu hết các thư viện đều đưa các mục lục thư viện của mình lên Web để người sử dụng truy cập, tìm tin và khai thác kho tài liệu của thư viện của mình.

3. TÌM TIN TRÊN WEB

Internet là một kho dữ liệu khổng lồ, đa dạng, không có khổ mẫu thống nhất, không được sắp xếp và quản lý theo một thể thức thống nhất. Vì thế để tìm được thông tin trên Internet đòi hỏi phải có những công cụ và phương pháp phù hợp. Do sự phát triển có tính chất bùng nổ của dịch vụ Web, nên đôi khi việc tìm tin trên Internet được coi là đồng nghĩa với tìm tin trên Web.

Phương pháp tìm tin là cách thức thực hiện việc tìm tin. Trên Web, có thể áp dụng hai phương pháp tìm như sau:

- Phương pháp tìm tin bằng từ khoá;
- Phương pháp tìm theo chủ đề.

Phương pháp tìm tin bằng từ khoá là cách thức tìm tin sử dụng những từ khoá và máy tìm tin để tìm ra những thông tin phù hợp với yêu cầu tìm kiếm. Tìm tin bằng từ khoá trên Web có thể coi như tương đương với việc tra cứu mục lục trực tuyến của một thư viện khổng lồ.

Phương pháp tìm theo chủ đề sử dụng những danh bạ chủ đề (đôi khi được gọi là danh bạ chủ đề) (subject directories) được xây dựng đặc biệt để quản lý nguồn tin trên Web và cung cấp khả năng truy cập đến chúng. Phương pháp tìm theo chủ đề có thể được coi là tương tự như sử dụng hộp phiếu mục lục thư viện.

Công cụ tìm tin trên Internet bao gồm một số loại cơ bản. Dựa vào phương pháp tìm tin, người ta có thể chia các công cụ tìm tin trên Web thành các nhóm:

- Máy tìm tin (Search engine): sử dụng để tìm tin bằng từ khoá.
- Danh bạ chủ đề (Subject directory): sử dụng phương pháp liệt kê theo chủ đề.
- Hướng dẫn chuyên môn hoá (Subject guide): sử dụng phương pháp liệt kê để tìm tin.
- Cơ sở dữ liệu.

Hiện nay, nhiều danh bạ chủ đề đều có các máy tìm tin riêng cho cơ sở dữ liệu về trang Web của mình. Ngược lại, những máy tìm tin cũng có danh bạ chủ đề đơn giản giúp người dùng tin tra cứu một số thông tin theo phương pháp liệt kê.

3.1. Máy tìm tin

Máy tìm tin là một chương trình tìm cho phép người sử dụng dùng các từ khoá (keyword) để tìm kiếm trong những cơ sở dữ liệu về trang web. Máy tìm tin bao gồm một cơ sở dữ liệu chứa các chỉ dẫn về hàng triệu trang web trên Internet và công cụ tìm kiếm. Người tìm tin giao tiếp với cơ sở dữ liệu, gửi yêu cầu tìm dưới dạng các từ khoá để "hỏi" cơ sở dữ liệu. Kết quả của lệnh tìm sẽ là những trang web có chứa các siêu liên kết tới những trang web thoả mãn yêu cầu của lệnh tìm. Có hàng trăm máy tìm tin như vậy trên toàn mạng Web. Để sử dụng được máy tìm tin, người sử dụng phải biết địa chỉ URL của máy và để truy cập trang Web đó bằng trình duyệt Web.

Máy tìm tin có thể được chia thành hai loại:

- Máy tìm tin đơn

- Máy siêu tìm tin

3.1.1. Máy tìm tin đơn

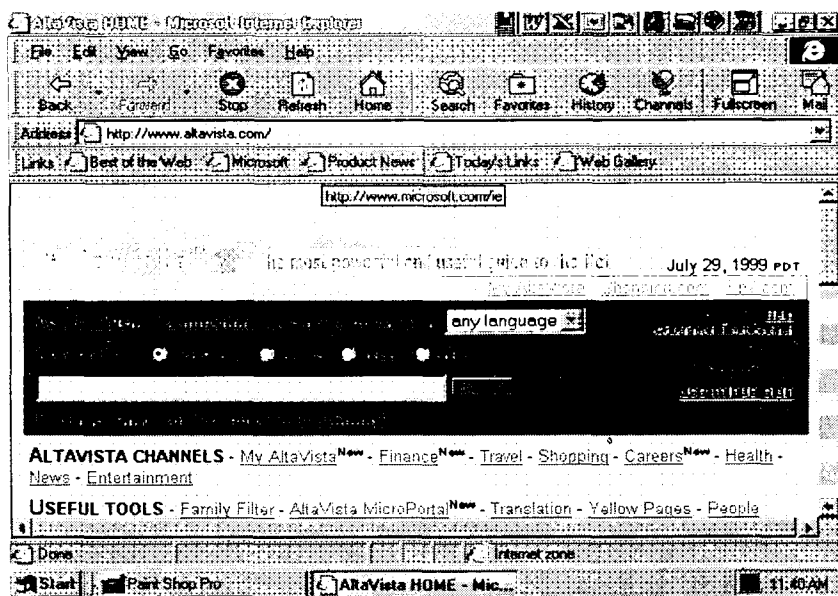
Trên Web có hàng trăm máy tìm tin. Đôi khi máy tìm tin còn được gọi là động cơ tìm kiếm. Máy tìm tin cung cấp một giao diện tương tác giữa người tìm tin và cơ sở dữ liệu của máy tìm tin. Đặc điểm cơ bản của máy tìm tin đơn là việc tìm tin được thực hiện với cơ sở dữ liệu về trang Web. Phần lớn các cơ sở dữ liệu của các máy tìm tin về trang Web thường được tạo ra một cách tự động bởi những chương trình đặc biệt. Các chương trình này được gọi bằng những tên khác nhau như "con nhện" (spiders), "người máy" (robots), Crawler,... Đó là những chương trình máy tính được thiết kế đặc biệt để thu thập thông tin về các trang Web và tạo ra cơ sở dữ liệu về chúng. Mỗi máy tìm tin đều có biểu mẫu (FORM) để nhận yêu cầu tin. Yêu cầu tìm là những từ khoá, cụm từ, hoặc những điều kiện tìm kiếm mà máy tìm tin cung cấp.

Dựa trên một sự kết hợp giữa các từ khoá, máy tìm tin sẽ tìm thông tin về tài liệu Web từ cơ sở dữ liệu về trang Web. Cần lưu ý rằng khi sử dụng máy tìm tin, chúng ta không tìm kiếm trực tiếp những thông tin trong Internet như nó đang tồn tại ở thời điểm đó, mà tìm kiếm những cơ sở dữ liệu cố định. Chúng đã được tạo ra từ trước (bởi các "con nhện" hay "người máy").

Đặc điểm chung của các máy tìm tin là việc định chỉ số cho các trang Web để đưa vào cơ sở dữ liệu được thực hiện tự động dựa trên toàn văn tài liệu HTML, không có phân loại theo chủ đề.

Hiện trên Web có hàng trăm máy tìm tin khác nhau. Do mỗi máy thực hiện công việc tìm tin theo các cách khác nhau nên cùng một yêu cầu tin, các máy tìm tin khác nhau sẽ đưa ra những kết quả khác nhau. Các yếu tố ảnh hưởng đến kết quả tìm được bao gồm:

- Kích thước (độ lớn) của cơ sở dữ liệu;
- Tính thường xuyên cập nhật thông tin;
- Các khả năng tìm kiếm.



Hình 13. Trang chủ máy tìm tin AltaVista

Những máy tìm tin cũng khác nhau trong tốc độ tìm kiếm, trong cách thiết lập giao diện tìm kiếm, cách hiển thị kết quả và khối lượng hướng dẫn (help) đưa ra.

Ưu điểm cơ bản của máy tìm tin là mức độ cập nhật nhanh hơn danh bạ chủ để do sử dụng các công cụ tự động trong việc xây dựng cơ sở dữ liệu. Do sử dụng công cụ tự động, các cơ sở dữ liệu về trang Web của máy tìm tin thường lớn (hàng chục triệu đến hàng trăm triệu biểu ghi), mức độ bao quát cao hơn, kết quả tìm kiếm thường đầy đủ hơn. Máy tìm tin được sử dụng hiệu quả hơn đối với các câu hỏi có tính đặc thù, chuyên sâu.

Máy tìm tin cũng có những nhược điểm. Do không có khả năng kiểm soát trở lại các liên kết nên nhiều trang Web đã không còn trên mạng nhưng vẫn có trong cơ sở dữ liệu, vì thế đôi khi dẫn đến kết quả tìm không đúng với thực tế. Do việc sử dụng phương pháp định chỉ số tự động nên mức độ nhiễu cao. Kết quả tìm tin nhiều khi rất lớn (hàng triệu biểu ghi), nên gây khó khăn cho việc lựa chọn kết quả phù hợp.

Để giải quyết vấn đề đưa nhiều kết quả không phù hợp, các máy tìm tin thường cung cấp khả năng xếp hạng biểu ghi theo mức độ phù hợp với yêu cầu tìm. Thông thường những biểu ghi có độ phù hợp cao hơn được trình bày đầu tiên. Tiêu chí xếp hạng không hoàn toàn như nhau giữa các máy tìm tin.

Có thể nêu một số ví dụ về máy tìm tin như sau:

- VinaSeek (<http://www.vinaseek.com>)
- PanVietnam (<http://www.panvietnam.com/>)
- Google (<http://www.google.com/>)
- AltaVista (<http://altavista.digital.com>)
- Excite (<http://www.excite.com>)
- HotBot (<http://www.hotbot.com>)

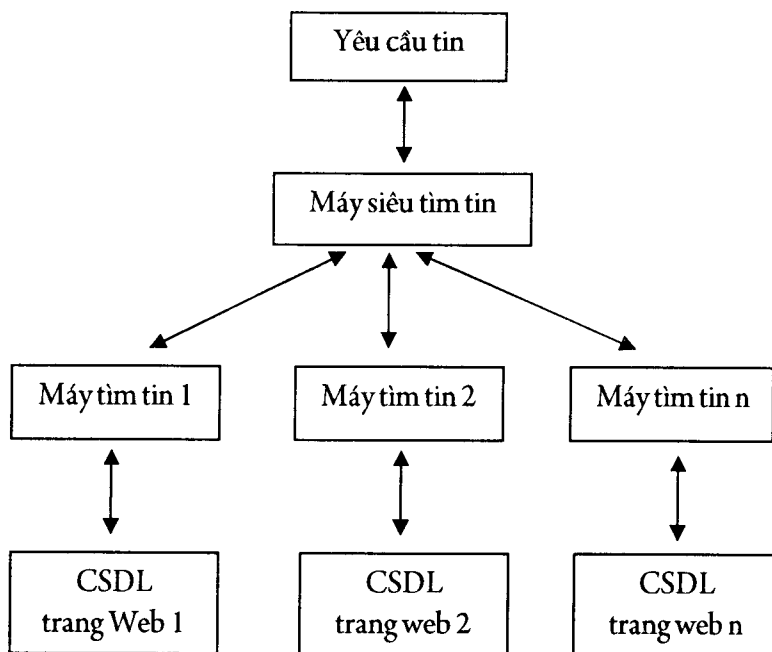
- Infoseek (<http://www.infoseek.com>)
- Northern Light (<http://www.northernlight.com>)

3.1.2. Máy siêu tìm tin

Do số lượng các máy tìm tin tăng nhanh và cách thức tìm kiếm cũng đa dạng, đã dẫn tới sự ra đời của những "siêu" công cụ tìm kiếm ("meta" search tools), được gọi là những **máy siêu tìm tin** (metasearch engines) hay máy tìm tin đa mạch (Multi-threaded search engine). Máy siêu tìm là một giao diện đặc biệt cho phép người tìm chỉ một lần nhập từ khoá có thể tìm kiếm đồng thời cùng lúc nhiều loại cơ sở dữ liệu của một số máy tìm tin khác nhau.

Điểm khác biệt cơ bản về cơ chế tìm tin giữa máy tìm tin đơn và máy siêu tìm tin là máy siêu tìm tin không có cơ sở dữ liệu về trang Web của chính mình. Chúng chỉ là một giao diện, nhận yêu cầu tìm và gửi chúng tới những máy tìm tin bình thường có cơ sở dữ liệu riêng để xử lý và nhận lại kết quả. Hình 14 giới thiệu quan hệ giữa máy tìm tin đơn và máy siêu tìm tin.

Nhược điểm của máy siêu tìm tin là chúng không có được mức độ kiểm soát hay khả năng kết hợp cao trong việc xây dựng biểu thức tìm như với giao diện tìm kiếm và logic tìm kiếm của các máy tìm tin đơn. Tuy nhiên, chúng có ưu điểm là những phần lớn máy siêu tìm tin chạy rất nhanh, có một số tính năng như khả năng phân loại sắp xếp kết quả theo địa chỉ (site), theo loại nguồn tin, theo miền (domain), khả năng lựa chọn các máy tìm tin, khả năng sửa đổi các kết quả, loại bỏ trùng lặp.



Hình 14. Quan hệ giữa máy siêu tìm tin và máy tìm tin đơn

Ví dụ về một số máy siêu tìm tin như:

- Dogpile (<http://www.dogpile.com>)
- Inference Find (<http://www.inference.com/infind/>)
- Meta Crawler (<http://www.metacrawler.com>)
- Profusion (<http://prufusion.ittc.ukans.edu/>)

3.1.3. Chế độ tìm tin của máy tìm tin

Thông thường các máy tìm tin có hai chế độ tìm: tìm cơ bản (hay tìm đơn giản) và tìm nâng cao (tìm phức hợp). Thuật ngữ chỉ các chế độ tìm có thể khác nhau với những máy tìm tin khác nhau.

Trong chế độ tìm cơ bản, người tìm không sử dụng toán tử mà thường chỉ sử dụng các từ riêng rẽ, tên riêng hoặc cụm từ. Các từ được viết vào cửa sổ câu hỏi sẽ được kết hợp theo phương thức ngầm định. Tùy theo máy tìm tin, phương thức kết hợp ngầm định có thể là AND hay là OR.

Để tìm chính xác theo cụm từ người dùng nên sử dụng dấu ngoặc kép “ ”. Điều này sẽ giúp giảm số trang tìm được, nâng cao tính chính xác của kết quả tìm tin (xem phần toán tử).

Để đạt kết quả tìm chính xác, đặc thù phải sử dụng tính năng tìm nâng cao. Mỗi máy tìm khác nhau có thể có các tên khác nhau cho chế độ tìm nâng cao: Advanced Search (với AltaVista), Power Search (với Excite), Super Search (với HotBot)...

Trong tìm nâng cao phải sử dụng các toán tử, các tiền tố để xác định tìm giới hạn trong trường.

3.1.4. Phương thức tìm tin

Tương tự như tìm tin tự động hoá, các máy tìm tin có thể sử dụng những phương thức tìm tin khác nhau. Có thể phân biệt một số phương thức chính sau: Toán tử boole (AND, OR, NOT); Dấu cộng/ trừ (+ hoặc -); Cụm từ trong dấu ngoặc kép “ ”; Chặt cụt (Truncation); Phân biệt chữ hoa/ chữ thường; Tìm hạn chế theo trường.

3.1.4.1. Toán tử boole

Toán tử boole là các quy tắc hoặc các lệnh đặc biệt sử dụng để tạo sự kết hợp giữa các từ khoá trong câu hỏi nhằm đạt được sự tìm kiếm có kết quả chính xác và đúng yêu cầu hơn. Tương tự như tìm tin tự động hoá, các máy tìm tin có thể sử dụng những ký hiệu khác nhau để thể hiện các toán tử boole chính AND, OR, NOT và NEAR.

- AND** Toán tử yêu cầu hai từ phải cùng có mặt trong tài liệu.
- OR** Toán tử đòi hỏi phải có ít nhất một trong hai từ phải có trong tài liệu.
- NOT** Toán tử yêu cầu từ đứng sau toán tử phải không có trong tài liệu.
- NEAR** Toán tử yêu cầu hai thuật ngữ phải gần nhau hoặc cách nhau một số từ nhất định trong biểu ghi.

Một điểm cần lưu ý là mỗi máy tìm tin đều có toán tử ngầm định riêng. Đa số các máy tìm tin sử dụng AND là toán tử ngầm định khi viết hai thuật ngữ tìm sát nhau.

Một số máy tìm tin còn sử dụng cả toán tử lân cận NEAR. Khi có toán tử NEAR, máy tìm tin sẽ chỉ tìm ra những tài liệu có hai từ trong biểu thức tìm cách nhau một số từ nhất định.

3.1.4.2. Dấu cộng (+)/dấu trừ (-)

Dấu cộng (+) và dấu trừ (-) là những toán tử thông dụng của nhiều máy tìm tin trên Internet.

Khi đặt dấu cộng (+) trước một từ trong biểu thức tìm, thì máy tìm tin sẽ yêu cầu tài liệu tìm được phải có chứa từ đó. Toán tử này có tác dụng tương đương toán tử AND. Chú ý không được đặt khoảng trống giữa dấu cộng (+) và từ. Ví dụ:

information +technology

Khi đặt dấu trừ (-) ở trước một từ thì máy tìm tin sẽ yêu cầu tài liệu tìm được phải không chứa từ đó. Toán tử này có tác dụng tương đương toán tử NOT. Chú ý rằng không được đặt khoảng trống giữa dấu trừ (-) và từ. Ví dụ:

information -technology

3.1.4.3. Dấu ngoặc kép " "

Khi những từ được đặt giữa hai dấu nháy kép thì chúng sẽ được xử lý như một cụm từ chính xác. Khi viết một biểu thức tìm có nhiều từ, nếu không có dấu ngoặc kép, máy tìm tin có thể áp dụng kết hợp ngầm định giữa các từ đó. Vì vậy để đảm bảo tìm chính xác hơn, cần sử dụng dấu nháy kép. Ví dụ:

"search tutorial for beginners"

3.1.4.4. Dấu chặt cụt (Truncation)

Kỹ thuật sử dụng dấu chặt cụt để mở rộng kết quả tìm tin cũng được áp dụng đối với nhiều máy tìm tin tương tự như trong tìm tin tự động hoá.

Chặt cụt là kỹ thuật mà ở đó người tìm tin sử dụng một ký hiệu chặt cụt được quy định cho máy tìm tin đặt phía cuối của thuật ngữ của thuật ngữ tìm làm cho hệ thống sẽ tìm ra những biểu ghi có chứa những thuật ngữ có phần gốc của từ trùng với phần ở trước ký hiệu chặt cụt.

Dấu chặt cụt rất có tác dụng trong việc tìm ra những từ có cách viết khác nhau (color, colour) hoặc những từ có chung gốc (có cấu tạo gốc giống nhau, ví dụ: psychology, psychological, psychologist, psychologists ...).

Ở những máy tìm in khác nhau, ký tự chặt cụt được quy định khác nhau. Nó có thể là: *, #, ?. Một vài máy tìm chỉ cho phép chặt phải (psycholog*), trong khi các máy khác lại có thể cho phép chặt giữa (colo*r).

Không phải tất cả các máy tìm tin đều hỗ trợ kỹ thuật chặt cụt này.

3.1.4.5. Phân biệt chữ hoa/ chữ thường

Do máy chủ nối mạng Internet sử dụng rất nhiều hệ điều hành mạng khác nhau nên có hiện tượng có sự phân biệt giữa dùng chữ thường và chữ viết hoa trong biểu thức tìm. Những hệ thống như vậy sẽ cho kết quả khác nhau khi dùng chữ thường và chữ in hoa làm câu hỏi tìm tin.

Trong những trường hợp có sự phân biệt chữ hoa và chữ thường, nguyên tắc xử lý chung của các máy tìm tin như sau:

- Nếu sử dụng chữ thường để tìm tin, thì những biểu ghi có thuật ngữ như vậy nhưng viết hoa đều được chấp nhận.
- Nếu sử dụng chữ hoa để lập biểu thức tìm thì chỉ biểu ghi có chứa thuật ngữ được viết hoa như trong yêu cầu tìm mới thoả mãn yêu cầu tìm.

Ví dụ, nếu chúng ta sử dụng máy tìm tin AltaVista để tìm tin, kết quả tìm bằng biểu thức "information technology" sẽ khác với kết quả tìm của biểu thức tìm "INFORMATION TECHNOLOGY" hoặc "Information Technology". Kết quả tìm tin của biểu thức "information technology" sẽ bao quát cả kết quả của biểu thức tìm tin "INFORMATION TECNOLOGY", nhưng không phải là ngược lại.

Sử dụng chữ hoa khi viết biểu thức tìm có hiệu quả cao khi tìm theo tên người, tên cơ quan/tổ chức,...

3.1.4.6. Tìm giới hạn theo trường

Mặc dù các trang Web không được cấu trúc hoá như những biểu ghi của cơ sở dữ liệu, để nâng cao hiệu quả tìm tin, đảm bảo thực hiện được một số yêu cầu đặc biệt, các máy tìm tin cũng cung cấp khả năng tìm tin hạn chế theo trường.

Phương pháp tìm giới hạn theo trường thường là sử dụng một số tiền tố đặt trước từ khoá tìm tin. Quy tắc chung khi viết yêu cầu tìm theo trường như sau:

Tiền tố: từ khoá

trong đó:

- Tiền tố là mã gán trước từ khoá xác định sẽ tìm giới hạn theo trường đó.

- Từ khoá: là từ khoá sử dụng để tìm tin.

Với những máy tìm tin khác nhau thì tiền tố cũng khác nhau. Những trường được sử dụng trong tìm giới hạn theo trường có thể khác nhau trong từng máy tìm tin.

Ví dụ máy tìm tin AltaVista cho phép tìm giới hạn theo một số trường như trình bày trong bảng.

Bảng 12. Tiền tố sử dụng trong tìm theo trường của AltaVista

Tiền tố	Ví dụ lệnh tìm	Ý nghĩa
anchor:	anchor:AltaVista	Tìm trang WEB có từ khoá cụm từ nằm trong phần văn bản của siêu liên kết. (ví dụ <u>Click here to Visit AltaVista</u>)
domain:	domain:vn domain:gov.vn	Tìm các trang WEB nằm trong một miền (domain). (tìm trang web có domain name chứa cụm từ vn, gov.vn).

host:	host:altavista.digital.com	Tìm các trang WEB nằm trong một máy chủ nào đó. (tìm trang web chứa trên máy chủ altavista.digital.com).
link:	link:anu.gov.au	Tìm các trang WEB có liên kết đến một trang WEB được ghi trong phần văn bản sau từ khoá LINK.
text:	text:NACESTID	Tìm các trang WEB có chứa từ khoá trong phần thân văn bản.
title:	title:Vietnam	Tìm trang WEB có từ khoá trong câu hỏi nằm trong vùng nhan đề.
url:	url:gov.vn	Tìm các trang WEB có những từ trong câu hỏi nằm trong vùng địa chỉ URL.

Máy tìm tin Vinaseek của Việt Nam (<http://www.vinaseek.com>) hoặc PanVietnam (<http://www.panvietnam.com>) cũng hỗ trợ tìm kiếm giới hạn theo trường. Ví dụ về các tiền tố dùng trong tìm kiếm giới hạn trong trường của Vinaseek trình bày trong bảng 13.

Bảng 13. Các tiền tố dùng trong tìm kiếm của Vinaseek

Tiền tố	Miêu tả	Ví dụ
	Nếu không để rõ trường tìm kiếm, mặc định sẽ là tìm trong toàn bộ nội dung văn bản, bao gồm cả tiêu đề và tóm tắt.	"thị trường chứng khoán"
link:	Có chứa liên kết siêu văn bản bên trong nội dung văn bản	link:tinhvan.com.vn/vns
site:	Tìm các văn bản trên máy chủ có địa chỉ site	site:vinaseek.com
url:	URL của văn bản tìm được	url:ww.tinhvan.com/help
title:	Tìm trong tiêu đề của văn bản	title:"bản tin nhanh"

Tổng hợp một số tính chất hỗ trợ tìm tin của một số máy tìm tin được trình bày trong bảng 14.

Tìm tin bằng từ khoá thường thực hiện khi có yêu cầu tin đặc thù, không chung chung. Để tìm tin hiệu quả, cũng cần có chiến lược phù hợp.

- Tìm hiểu sơ bộ máy tìm tin;
- Tìm hiểu yêu cầu/câu hỏi;
- Xác định các khái niệm quan trọng của yêu cầu;
- Xác định từ khoá để mô tả khái niệm;

- Nhập yêu cầu tìm. Lúc đầu không nên sử dụng các kết hợp phức tạp mà chỉ dùng 2-3 từ khoá;
- Sử dụng dấu "..." để thể hiện những cụm từ;
- Lúc đầu nên sử dụng chữ thường vì cho kết quả cao hơn. Nhưng nếu biết chính xác thì nên sử dụng chữ hoa vì cho kết quả chính xác hơn (next, neXt,...).

• Để nâng cao tốc độ tìm tin theo từ khoá nên đánh dấu lưu địa chỉ các máy tìm tin (dùng FAVORITES của Explorer hoặc BOOKMARK của Netscape).

Thường xuyên xem lướt hướng dẫn (HELP) vì các máy tìm tin được cải tiến liên tục.

Bảng 14. Tính năng của một số máy tìm tin

Toán tử Máy tìm tin	boole	+/-	Dấu ngoặc kép ""	Chặt cụt	Phân biệt chữ hoa/thường
<i>Google</i>	x	x	x	x	
<i>AltaVista</i>	x	x	x	x	
<i>Excite</i>	x	x	x	x	x
<i>Hotbot</i>	x	x	x		x
<i>Infoseek</i>		x	x		x
<i>DogPile</i>	x		x		

Ghi chú: x: là có tính năng đó.

bỏ trống là không có khả năng đó.

3.2. Danh bạ chủ đề

Ngoài phương pháp tìm kiếm dùng từ khoá, trên Web còn có một loại công cụ khác để tìm tin gọi là Danh bạ chủ đề (tiếng Anh là Subject Directory). Danh bạ chủ đề là một chi mục có tổ chức về các nguồn tin có trên Web. Có thể so sánh Danh bạ chủ đề với các bảng mục lục chủ đề của thư viện. Đây là phương pháp tìm tin theo kiểu liệt kê, truy cập theo các đề mục chủ đề có tính thứ bậc. Thông thường việc tìm kiếm bắt đầu bằng một đề mục chủ đề chung nhất, sau đó đi sâu (thu hẹp) dần đến những chủ đề hẹp và chuyên hơn. Vì vậy có thể coi đây là phương pháp tìm theo chủ đề.

Việc sắp xếp các chủ đề theo thứ bậc, cho phép người sử dụng biết được vị trí các chủ đề trong Web nhằm tìm kiếm những thông tin cần thiết (xem hình 15). Hình 15 là ví dụ về một số lớp đề mục chủ đề của Yahoo. Khác với các cơ sở dữ liệu về trang Web của máy tìm tin, các danh bạ chủ đề được tạo ra và duy trì bởi con người.

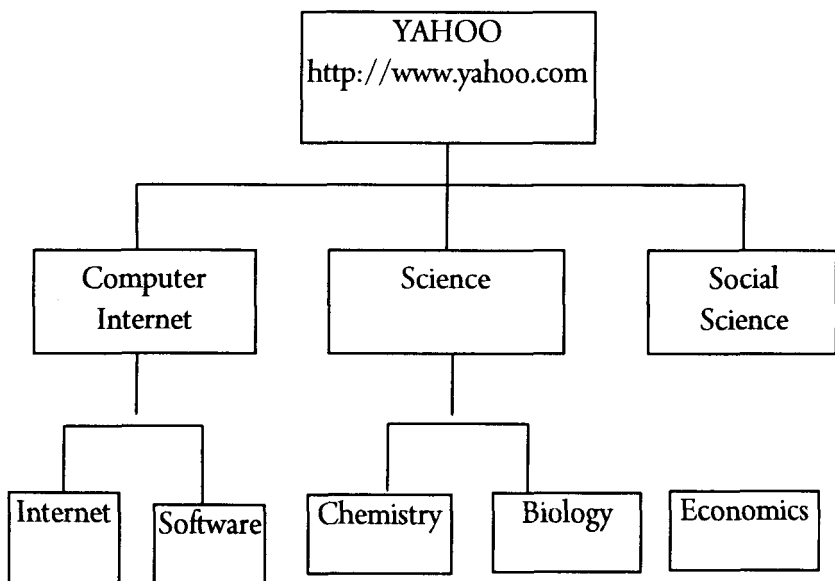
Khác với tình trạng về hệ thống phân loại sử dụng trong thư viện (chỉ có một số hệ thống phân loại), không sự thống nhất về hệ thống phân loại chủ đề trên Web. Mỗi danh bạ chủ đề sử dụng hệ thống phân loại riêng của mình. Điều này gây khó khăn cho việc sử dụng, bởi người dùng tin phải tìm hiểu kỹ từng danh mục mới có thể khai thác chúng hiệu quả.

Ngoài ra, do các danh bạ chủ đề được xây dựng và duy trì bởi con người nên chúng thường có quy mô nhỏ hơn cơ sở dữ liệu của máy tìm tin, không bao quát được toàn bộ Web, kết quả tìm thường ít hơn.

Tuy nhiên, do danh mục được xây dựng bởi chuyên gia, được chọn lọc và đánh giá cẩn thận về tính phù hợp với chủ đề nên tuy

kết quả tìm đưa ra thấp hơn, song mức độ phù hợp với yêu cầu cao hơn. Hơn nữa, do đặc trưng chịu ảnh hưởng của người sử dụng, danh bạ chủ đề giảm một cách đáng kể việc lấy ra những kết quả không cần thiết.

Thông thường máy tìm tin lập chỉ mục tất cả các trang Web của một Web site nhất định trong khi danh bạ chủ đề thường chỉ tạo ra một mối liên hệ duy nhất với trang chủ (home page) của Website đó.



Hình 15. Ví dụ về sơ đồ phân cấp một số chủ đề của Yahoo

Để tìm tin theo danh bạ chủ đề, người tìm phải truy cập trang chủ của danh mục chủ đề đó. Tìm kiếm trên Web bằng danh bạ chủ đề có ưu điểm là thông tin do được sắp xếp vào chủ đề bởi các chuyên gia nên kết quả đưa ra thường chính xác. Tuy nhiên, nó có

nhược điểm là độ lớn của cơ sở dữ liệu thường bị hạn chế (nhỏ) và cập nhật chậm hơn.

Bởi vì các danh bạ chủ đề được sắp xếp theo phân loại và chúng dễ dàng quay lại những kết nối với cấp cao nhất (top - level) của một Web site hơn là với những trang riêng biệt nên chúng phù hợp với việc tìm kiếm thông tin về một chủ đề chung chung hơn là một mảng thông tin cụ thể.

Rất nhiều trong số danh bạ chủ đề có một máy tìm tin riêng phục vụ cho việc tìm kiếm cơ sở dữ liệu riêng. Các cơ sở dữ liệu của đề mục chủ đề thường nhỏ hơn cơ sở dữ liệu của các máy tìm tin, tức là những danh sách kết quả đưa ra cũng nhỏ hơn.

Một số ví dụ về danh bạ chủ đề bao gồm:

Yahoo (<http://www.yahoo.com/>)

Galaxy (<http://www.einet.net>)

Look Smart (<http://www.looksmart.com>)

Magellan ()

3.3. Hướng dẫn chủ đề chuyên môn hoá

Do Web chứa một khối lượng thông tin khổng lồ nhưng vô cùng đa dạng, khác nhau về lĩnh vực và chủ đề, cộng thêm sự phát triển rất nhanh chóng và thay đổi thường xuyên, việc nắm bắt một cách cập nhật những thông tin quan trọng trong mọi chủ đề là điều không thể làm được với một vài người. Do đó những danh bạ chủ đề chung như Yahoo chỉ có thể bao quát những lĩnh vực chung nhất mà thôi, trong khi các chuyên gia có thể cần thông tin đầy đủ hơn về lĩnh vực hẹp mà họ quan tâm. Để giải quyết vấn đề này, người ta tạo những **Hướng**

dẫn chủ đề chuyên môn hoá (Specialized Subject Guide). Đó là một hướng dẫn được soạn ra bởi một số chuyên gia trong lĩnh vực cụ thể để tiếp cận những nguồn tài nguyên quan trọng thuộc lĩnh vực chuyên môn mà họ quan tâm đặc biệt.

Những Hướng dẫn chủ đề chuyên môn hoá là những trang Web chứa tập hợp những liên kết tới những thông tin cần thiết hiệu quả hơn về một chủ đề hẹp so với danh bạ chủ đề chung chung. Những hướng dẫn như vậy có thể tồn tại trong hầu hết mọi chủ đề.

Tương tự như những máy siêu tìm tin cho phép tìm đồng thời một số máy tìm tin khác nhau, một vài Web site đóng vai trò như những "Thư viện" hoặc những "Trung tâm chỉ dẫn" về các Hướng dẫn chủ đề chuyên môn hoá. Các site này thực hiện việc xem xét, kiểm tra lại cũng như đưa ra những chú giải cho các danh bạ chủ đề mà nó chứa đựng, và phần lớn công việc dựa trên nguyên tắc cho phép các chuyên gia chủ đề duy trì các danh bạ chủ đề chuyên môn hoá cá nhân. Một vài "Trung tâm chỉ dẫn" như vậy xây dựng các hướng dẫn chủ đề chuyên môn hoá ngay tại Web site của nó, trong khi một số khác lại chỉ chỉ dẫn các liên kết tới những hướng dẫn đặt tại nhiều site khác ở xa.

Một số ví dụ về điểm chỉ dẫn như vậy có thể là:

- Argus Clearinghouse (<http://www.clearinghouse.net>)
- The Mining Company (<http://www.miningco.com>)
- WWW Virtual Library ()

3.4. Cơ sở dữ liệu

Trên Web có hàng trăm cơ sở dữ liệu khác được các tổ chức, cơ quan cung cấp lên Web để truy cập. Các CSDL này được các tổ chức

xây dựng cho mục đích nội bộ nhưng được họ cung cấp lên mạng để người sử dụng Internet khai thác.

Các cơ sở dữ liệu này có thể bao quát các lĩnh vực khác nhau. Những Website cơ sở dữ liệu này cung cấp các công cụ tìm kiếm bằng từ khoá để người sử dụng Internet truy cập và tìm kiếm trong cơ sở dữ liệu, trong một máy chủ nào đó. Mỗi cơ sở dữ liệu có máy tìm tin riêng và cú pháp riêng vì thế để khai thác hiệu quả phải nghiên cứu kỹ phương pháp tìm kiếm của từng cơ sở dữ liệu.

Phương pháp tìm tin trên các cơ sở dữ liệu này nhìn chung tương tự như tìm tin tự động hoá và tìm tin bằng máy tìm tin.

HƯỚNG DẪN TỰ HỌC

1. Tài liệu cần đọc

1. Gosling, Mary. Learn reference work.- Canberra, DocMatr Pty Ltd, 1999.- 154p.
2. Grogan, Chris. Practical reference work.- London, Library Association, 1992.- 161p.
3. Kats William. A introduction to Reference work :
Vol. I: Basic Information Sources. 7th ed.- NY. McGraw-Hill, Inc, 1997.- 398 p
Vol. II: Reference services and reference processes. 7th ed.- NY. McGraw-Hill, Inc, 1997.- 259p.
4. Meadow, Charles T. Text Information Retrieval Systems.- Toronto, Academic Press, Inc, 1992.- 302p.
5. Online Information Retrieval : an Introductory Manual to Principles and Practice / John Convey.- Fourth edition.- London: Library Association, 1992. 310p.

2. Câu hỏi ôn tập

2.1. Chương 1:

Câu 1. Tra cứu thông tin là gì? Cơ sở phân loại các loại tra cứu?

Câu 2. Trình bày các yếu tố của hệ thống tra cứu thông tin?

Câu 3. Trình bày các loại ngôn ngữ tìm tin chính trong các cơ quan thông tin thư viện? Phân biệt sự giống và khác nhau của các loại ngôn ngữ tìm tin đó (Về cấu trúc; Về nội dung; Về mức độ ứng dụng...)?

Câu 4. Trình bày thành phần và công dụng của bộ máy tra cứu?

Câu 5. Trình bày vai trò của cán bộ tra cứu và NDT trong hệ thống tra cứu thông tin?

2.2. Chương 2:

Câu 1. Trình bày các nguyên tắc tra cứu?

Câu 2. Nêu những yếu tố chính để thực hiện việc tra cứu?

Câu 3. Trình bày các bước của quá trình tra cứu?

Câu 4. Nêu cơ sở đánh giá hiệu quả tra cứu? Các công thức tính? Mục đích đánh giá?

2.3. Chương 3:

Câu 1: Tra cứu thông tin truyền thống là gì? Các yếu tố chính để tiến hành tra tìm?

Câu 2: Trình bày công dụng của các nguồn tra cứu thông tin tư liệu truyền thống?

Câu 3: Trình bày phương pháp tra cứu thông tin tư liệu qua hệ thống mục lục và ấn phẩm thông tin?

Câu 4: Tra cứu thông tin dữ kiện là gì? So sánh sự khác nhau và giống nhau của tra cứu thông tin tư liệu và thông tin dữ kiện? Trình bày cơ sở để tra cứu thông tin dữ kiện.

Câu 5: Trình bày công dụng của các nguồn tra cứu thông tin dữ kiện truyền thống và những đặc thù của thông tin dữ kiện? phương pháp tra cứu thông tin dữ kiện?

2.4. Chương 4:

Câu 1. Biểu thức tìm tin là gì? Có những loại thuật ngữ tìm tin nào?

Câu 2. Hãy trình bày các toán tử boole trong tìm tin tự động hoá?

Câu 3. Hãy giới thiệu trình bày các toán tử tìm tin không thuộc nhóm toán tử boole?

Câu 4. Chiến lược tìm tin tự động hoá là gì? Có những chiến lược tìm tin cơ bản nào?

Câu 5. Tìm tin trực tuyến là gì? Hãy cho biết có những thành phần cơ bản nào tham gia vào công nghiệp thông tin trực tuyến?

Câu 6. CD-ROM là gì? Hãy so sánh tìm tin trên CD-ROM và tìm tin trực tuyến?

2.5. Chương 5:

Câu 1. Hãy cho biết Internet là gì? Trình bày các khái niệm địa chỉ IP, địa chỉ tên miền. Giới thiệu sơ lược những dịch vụ cơ bản của Internet.

Câu 2. World Wide Web là gì? Hãy cho biết có những nguồn tin gì trên Internet?.

Câu 3. Hãy giới thiệu phương pháp bảng từ khoá và công cụ tìm tin liên quan trên Internet?

Câu 4. Hãy giới thiệu phương pháp tìm tin theo chủ đề?

3. Bài tập thực hành

□ 3.1. Tìm tin truyền thống

a) Tra cứu thông tin thư mục

1. Tập thơ Từ ấy của nhà thơ Tố Hữu được tái bản bao nhiêu lần? Vào những năm nào?
2. Có bao nhiêu tác phẩm của dịch giả Đoàn Tử Huyền có trong thư viện trường ĐHVH? Nêu tên những tác phẩm đó?
3. Ai viết truyện ngắn *Những người thích đùa*?
4. Lập danh mục tài liệu về mổ nội soi tiêu hoá xuất bản từ năm 2000-2003 ?
5. Hai nhà thơ Xuân Diệu và Huy Cận xuất bản bao nhiêu tập thơ chung? Nêu tên và năm xuất bản những tập thơ đó?
6. Nêu mã ngành và nơi bảo vệ luận án tiến sỹ (LATS) của NCS Nguyễn Văn Cẩn?
7. Tìm các tài liệu có đề cập tới Sông Mã xuất bản từ năm 2001 - 2002?
8. Nêu tên các tác giả có nhiều công trình đề cập đến Bảo vệ động vật quý hiếm ở VN?
9. Tìm 5 tài liệu bằng tiếng Khmer xuất bản từ năm 1999 - 2002?
10. Cho biết tên các bài báo tạp chí mới nhất đề cập tới ô nhiễm nguồn nước thành phố?
11. Aizmatov là nhà văn nước nào? Có tác phẩm nào nổi tiếng? Những tác phẩm nào đã được dịch sang tiếng Việt?

b) Tìm tin dữ kiện

1. Cho biết năm sinh và năm mất của Hưng Đạo Vương Trần Quốc Tuấn?

2. Nữ hoàng Elizabeth II đăng quang khi nào?
3. Nicolas Copernic là ai? Tóm tắt tiểu sử của ông?
4. Tóm tắt lịch sử kênh đào Panama?
5. Ngày 9 tháng 5 năm 1945 là ngày gì?
6. Năm 1898 ai là Vua của Việt Nam?
7. Cho biết trụ sở và điện thoại giao dịch của Ngân hàng Ngoại thương VN?
8. Chức năng của Tổ chức Y tế Thế giới là gì?
9. Cho biết số liệu về sinh sản, kết hôn và chết ở Việt Nam từ năm 2000 - 2003.
10. Hiện nay ở VN có bao nhiêu Km đường quốc lộ?
11. Nêu tên ba chiếc cầu dài nhất thế giới hiện nay?
12. Người Tajikisstan nói tiếng gì?

□ 3.2. Tìm tin Tự động hoá

1. THỰC HÀNH TÌM TIN BẰNG CHƯƠNG TRÌNH WINISIS

Mục đích

Winisis là một chương trình lưu trữ và tìm tin rất phổ biến ở Việt Nam. Chương trình cung cấp các phương tiện tra cứu tin từ cơ sở dữ liệu rất tiêu chuẩn. Vì vậy việc học sử dụng Winisis sẽ cung cấp cho học viên những kỹ năng về sử dụng các toán tử, kết hợp toán tử để tạo câu hỏi phức tạp, sử dụng từ điển (tệp đảo của cơ sở dữ liệu để tra cứu tin).

Mục đích cụ thể của bài tập thực hành này là giúp học viên:

- Biết cách dùng lệnh tìm tin;

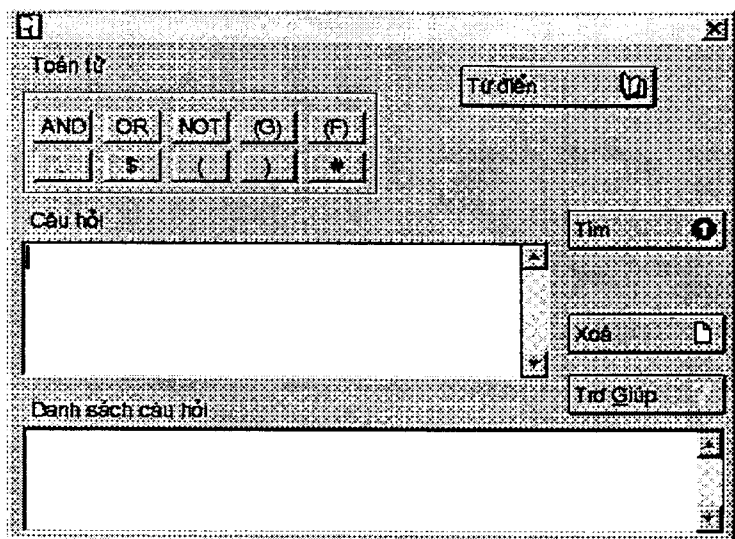
- Biết cách sử dụng toán tử: toán tử boole; toán tử lân cận; sử dụng ký hiệu chặt cụt;
- Tìm hiểu sự ưu tiên xử lý của toán tử;
- Biết sử dụng dấu ngoặc để tạo thay mức ưu tiên câu hỏi;
- Biết sử dụng toán hạng (/ID) để tìm hạn chế theo trường.

Khởi động chương trình

Sau khi khởi động chương trình Winisis, chúng ta thực hiện các bước chọn thực đơn như sau:

Tìm tin ==> Tìm tin trình độ cao

Chúng ta thấy cửa sổ tìm tin như sau:



Bài tập 1. Xem xét việc so sánh giữa "khái niệm" và "ký hiệu".

Bây này nhằm làm cho học viên hiểu rõ máy tính làm việc với ký hiệu mà không hiểu khái niệm.

Đánh các lệnh tìm sau và cho biết số biểu ghi tìm được:

thủy lợi

thuỷ lợi

hoá học

hóa học

Giải thích lý do khác nhau.

Bài 2. Sử dụng toán tử OR (dấu +) để mở rộng kết quả tìm.

Toán tử OR cho phép mở rộng kết quả tìm kiếm. Sử dụng với các từ đồng nghĩa hoặc mở rộng khái niệm. Tìm thử các câu hỏi:

thuỷ lợi + thủy lợi

hoá học + hóa học

Giải thích lý do số biểu ghi tăng lên so với tìm riêng "hoá học" hoặc "hóa học". Rút ra kết luận gì?

Bài tập 3. Sử dụng toán tử AND (dấu *)

Đánh các câu hỏi sau và cho biết số biểu ghi tìm được.

chọn giống lúa

chọn giống * lúa

đầu tư * nông nghiệp

đầu tư cho nông nghiệp

Giải thích sự khác nhau về kết quả tìm. Rút ra kết luận gì?

Bài tập 4. Sử dụng toán tử lân cận (dấu chấm).

Tìm tin bằng cụm từ và dùng toán tử lân cận, cho biết kết quả tìm được.

chọn giống * lúa

chọn . giống * lúa

nông nghiệp

nông . nghiệp

nông (f) nghiệp

Giải thích sự khác nhau về kết quả tìm. Rút ra kết luận gì?

Bài tập 5. Sử dụng ký hiệu chặt cụt (dấu \$).

Tìm tin bằng cụm từ và dùng toán tử lân cận.

chọn giống

chọn giống\$

thông tin

thông tin\$

Giải thích sự khác nhau về số biểu ghi tìm được. Rút ra kết luận gì?

Bài tập 6. Tìm hiểu mức ưu tiên của toán tử.

Đánh các biểu thức tìm tin như dưới đây và ghi lại thứ tự lệnh tìm và số biểu ghi tìm được.

$$A + B * C + D * E$$

$$(A + B) * (C + D) * E$$

$$A + B * (C + D * E)$$

Giải thích sự khác nhau trong số biểu ghi tìm được và thứ tự thực hiện lệnh.

Bài tập 7. Sử dụng toán hạng /(ID) để tìm hạn chế theo trường.

Thực hiện các biểu thức tìm sau và cho biết kết quả tìm tin.

chọn giống/(312)

chọn giống

lúa

lúa/(312)

lúa/(200)

chọn giống\$ * lúa/(200)

chọn giống\$ (F) lúa/(200)

Giải thích sự khác nhau trong số biểu ghi tìm được và thứ tự thực hiện lệnh.

Bài 8. Tự thực phân tích và thực hiện câu hỏi.

Hãy tìm tài liệu nói về ảnh hưởng của thuốc bảo vệ thực vật đối với thiên nhiên và con người.

2. THỰC TẬP VỚI CHƯƠNG TRÌNH DIALOG ONDISK

Mục đích:

Mục đích của những bài tập này nhằm:

- Làm quen với việc phân tích yêu cầu và chuẩn bị câu hỏi với hệ thống;
- Làm quen và nắm vững những lệnh cơ bản của chương trình

tìm tin của DIALOG.

- Thực tập áp dụng những chiến lược tìm tin khác nhau như: Xây dựng khối, Nuôi ngọc trai,...

Khởi động chương trình

Chương trình Dialog trên CD-ROM có tên tệp là ondisk.exe có trên máy chủ. Việc khởi động chương trình phụ thuộc vào cài đặt trên máy chủ.

Chọn và Kích đúp ONDISC.EXE

Khi chương trình khởi động chúng ta sẽ thấy thực đơn tương tự như sau:

Select Search Mode
Easy Menu Search
<i>Dialog Command Search</i>
Online Search
Setup and Accounting
Exit to DOS

Để thực hiện tìm tin theo thực đơn, ta chọn **Easy Menu Search**.

Để thực hiện tìm theo dòng lệnh, ta chuyển thành chọn lệnh đến:

Dialog Command Search

Hoặc bấm ký tự đại diện <D>

Chú ý: không dùng chuột mà phải dùng phím mũi tên để dịch chuyển đến chức năng cần chọn.

Chúng ta sẽ thấy thông báo như sau

Ei EEDisc(tm)

The Database for Computer and Electrical Engineering

(Ei) Engineering Information Inc.

*This database is a subset of Dialog's online File 8, Ei
COMPENDEX*PLUS(tm)*

More current records may be found online in File 8.

Copyright (c) 1994, (Ei) Engineering Information Inc. All rights reserved.

Copyright (c) 1994, DIALOG Information Services, Inc.

All rights reserved. No claim to original U.S. Gov't works.

Command Search

Dialog User Interface as Audio/Visual work,

Copyright (c) 1986 Dialog Information Services, Inc.

Portions (c) 1987, 1991 Digital Library Systems, Inc

Ei EEDisc 1992 - December 1994

Set Items Description

?

Tại dấu nhắc (?) ta đánh các lệnh.

Thí dụ:

?S computer

CÁC BÀI TẬP THỰC HÀNH

1. Lệnh tìm (Select) , lệnh xem kết quả (Type) và thoát ra (Logoff)

? **S information technology** Tìm các biểu ghi về công nghệ thông tin

- Tìm hiểu xem thông báo của hệ thống, nhận biết số của câu hỏi, số biểu ghi tìm được.

? **T S1/5/1-3** Xem kết quả tìm của câu hỏi số 1, theo format số 5 (đầy đủ), từ biểu 1-5

? **Logoff** Thoát ra khỏi lệnh tìm, về menu

2. Sử dụng lệnh trợ giúp tìm các format

Vào trở lại lệnh.

? **HELP** Hiện thông báo về các loại lệnh

? **HELP FORMAT** Hiện thông báo về các loại FORMAT.

Cho biết có bao nhiêu loại lệnh và có bao nhiêu loại Format ?

Liệt kê tên FORMAT

3. Sử dụng các toán tử AND, (W), (N) và so sánh kết quả.

Thực hiện lệnh tìm cho các câu hỏi sau và cho biết số biểu ghi tìm được.

? **S information technology**

? **S technology information**

? **S information AND technology**

? **S technology AND information**

? **S information (W) technology**

? **S technology (W) information**

? **S information (N) technology**

? **S technology (N) information**

? S information (3W) technology

? S information (3N) technology

? S information (L) technology

? S technology (L) information

? S information (S) technology

? S technology (S) information

Cho biết kết quả của từng lệnh tìm. Giải thích lý do của sự khác biệt đó.

4. Sử dụng toán tử chặt cụt ở cuối từ.

Thực hiện các lệnh tìm:

? S COMPUTER

? S COMPUTER?

? S COMPUT?

Cho biết mỗi biểu thức cho kết quả bao nhiêu biểu ghi. Giải thích sự khác nhau của kết quả.

5. Sử dụng toán tử chặt cụt (?) ở giữa từ.

Thực hiện lệnh tìm với các biểu thức sau:

? S WOMAN

? S WOMEN

? S WOM?N

? S organisation

? S organization

? S organi?ation

Cho biết mỗi lệnh cho bao nhiêu biểu ghi, biểu thức tìm nào cho số biểu ghi nhiều nhất?, Tại sao?.

6. Sử dụng EXPAND để xem từ điển và tìm tìm theo từ điển.

? EXPAND LIBRARY Xem từ điển bắt đầu bằng từ Library.

Chọn từ có số n để tìm tin

7. Sử dụng lệnh tìm kết hợp câu hỏi trước

? S LIBRARY Xem và nhớ số của câu hỏi (thí dụ là S1)

? S COMPUTER? Xem và nhớ số của câu hỏi (thí dụ là S2)

? S S1 AND S2

3.3. Tìm tin trên Internet

3. THỰC HÀNH TÌM TIN TRÊN INTERNET BẰNG YAHOO

Yêu cầu là phải có kết nối mạng Internet. Khi đã kết nối, chúng ta khởi động Internet Explorer và đánh vào ô ADDRESS địa chỉ URL sau:

<http://www.yahoo.com/>

1. Tìm theo liệt kê

a. Dùng phương pháp liệt kê tìm Website của Thư viện Quốc hội Mỹ.

Mở Yahoo ==> Chọn Reference ==> Libraries => National Libraries ==> United State Library of Congress.

b. Phương Liệt kê Dùng phương pháp liệt kê tìm Website Danh bạ công ty sản xuất ở Australia.

Mở Yahoo ==> Business and Economy ==> Directories => Companies ==> Manufacturing => Australian Manufacturer's Guide ==> xem công ty

c. Dùng Phương pháp liệt kê để tìm công cụ tra cứu

Mở Yahoo ==> Reference == > Acronyms and Abbreviation
==> Acronym Finder

Khi thấy cửa sổ chương trình tra cứu, đánh các từ viết tắt vào để tìm các từ viết tắt: KISS, MMDS, AIDS, SARS là gì, có mấy cách giải thích.

Vào một số mục và đi sâu dần vào theo chủ đề được chọn (Phương pháp liệt kê đơn giản).

2. Tìm theo từ khóa

Tại cửa sổ tìm tin, nhập yêu cầu cần tìm. Thí dụ:

Microsoft

Kích chuột nút SEARCH

Chọn mục cần/muốn xem trong kết quả thu được.

a. Thử nghiệm các toán tử:

Toán tử +/-

Xem sự khác nhau giữa kết quả của ba câu hỏi::

Vietnam War

Vietnam -War

Vietnam +War

Hoặc

Vietnam Culture Folklore

Vietnam Culture +Folklore

Vietnam -Culture +Folklore

b. Tìm giới hạn theo trường

t:IFLA

u:intel

c. Dấu "... "

Tìm hiểu sự khác nhau về kết quả giữa

"information science"

information science

và

"great barrier reef"

great barrier reef

4. THỰC HÀNH TÌM TIN TRÊN INTERNET VIỆT NAM BẰNG MÁY TÌM TIN VINASEEK

Cách thức tìm kiếm trên VinaSeek rất đơn giản: chỉ việc gõ vào ô tìm kiếm một số từ khoá hoặc câu tiếng Việt.

Để tìm tin bằng Vinaseek, thực hiện các bước sau:

- Truy cập trang Web của Vinaseek:

<http://www.vinaseek.com/>

- Chọn bộ mã ký tự muốn sử dụng: TCVN, UNICODE, VNI,...

- Nhập câu hỏi vào ô cửa sổ tìm kiếm.

Hãy nhập các câu hỏi sau, cho biết số trang tìm được, giải thích tại sao?

? **thị trường chứng khoán**

? **"thị trường chứng khoán"**

? "Thị Trường Chứng Khoán"

? +"thị trường" +"chứng khoán"

? +"thị trường chứng khoán"

? +"thị trường chứng khoán" +title:"bản tin nhanh"

Tài liệu tham khảo

I. Tài liệu tiếng Việt

1. Bảng phân loại. *Dùng cho các thư viện khoa học tổng hợp*. 2 T/ Thư viện Quốc gia Việt Nam.- H.: TVQG, 2002.
2. *Bảng phân loại Thư viện – Thư mục BBK/Thư viện Quốc gia Việt Nam*.- H.: TVQG, 2002.
3. Đảng CSVN. *Văn kiện đại hội đại biểu toàn quốc lần thứ IX*.- H.:CTQG., 2001.- 253tr.
4. Đoàn Phan Tân. *Cơ sở thông tin học* .-H.: Trường ĐHVH, 2001.- 139tr
5. Đoàn Phan Tân. *Tin học trong hoạt động thông tin - thư viện* .-H.: Văn hoá - Thông tin, 1997.- 242 tr.
6. *Khung Đề Mục hệ thống thông tin khoa học và kỹ thuật Quốc gia*.- H.: Viện thông tin Khoa Học và Kỹ thuật Trung ương, 1987.
7. *Khung phân loại Thập phân bách khoa*. H.: Viện thông tin Khoa Học và Kỹ thuật Trung ương, 1987.
8. *Khung phân loại Thư viện – Thư mục (BBK)*. 2T. - H.: Trung tâm thông tin khoa học và công nghệ QG, 2002.
9. *Kí yếu hội nghị ngành Thông tin tư liệu Khoa học và Công nghệ*. -H.: Trung tâm thông tin tư liệu khoa học và công nghệ VN, 1998.- 217 tr.

10. Lê Thị Chính. *Thu mục tài liệu KHKT.* - H.: Trường cao đẳng nghiệp vụ văn hoá, 1981.- 192 tr.
11. *Patrick Vincent. Internet toàn tập: Hướng dẫn - Thao tác - ứng dụng/* Đoàn Công Hùng: Dịch. - Tp. HCM: NXB Trẻ, 1997.
12. *Pfaffenberger, Bryan. Microsoft Internet Explorer 4.0 toàn tập/* Đoàn Công Hùng, Vũ Tài Hoa: Dịch; Nguyễn Văn Phước: H.dính.- NXB Trẻ, 1998.
13. Phan Thị Đém - Nguyễn Công . *Thu mục tài liệu Chính trị - Xã hội.* - H.: Trường cao đẳng nghiệp vụ văn hoá, 1981.- 238 tr.
14. Phan Văn. *Thông tin học.* -H.: Nxb Đại học QG HN, 2000.- 137 tr.
15. Srisakdi Charmonman et al. *Những cơ sở ban đầu về Internet/* Trần Hoàng Lương: dịch.- H.: Khoa học và Kỹ thuật, 1996. 214tr.
16. Tạ Thị Thịnh. *Phân loại tài liệu.* Giáo trình. - H.: Trung tâm thông tin tư liệu khoa học và công nghệ QG, 1998.
17. *Từ điển từ khoá Khoa học và Công nghệ.* 2T/ Trung tâm thông tin khoa học và công nghệ quốc gia. - H., 2001.
18. Trần Mạnh Tuấn. *Sản phẩm và dịch vụ thông tin thư viện.* - H.: Trung tâm thông tin tư liệu khoa học và công nghệ QG, 1998. - 324 tr.
19. Trịnh Kim Chi - Dương Bích Hồng. *Thu mục học đại cương.* - H.: Trường ĐHVH, 1993. - 246 tr.
20. Vũ Dương Thuý Ngà. *Định chủ đề tài liệu.* - H.: Văn hoá - Thông tin, 1995. - 133tr.
21. Vũ Dương Thuý Ngà. *Phân loại tài liệu.* - H.: Trường Đại học Văn hóa, Hà Nội, 2004. - 322tr.

II. Tài liệu nước ngoài

1. Alison Scanneti (edit). *Handbook of Special Librarianship and Information Works*. - 7th edition. - London: Aslib, 1997.
2. Chernyi, A.I. *Vvedeniye i teorii informacionogo poiska*. - M.: Nauka, 1975. (t.Nga).
3. Convey J. *Online Information Retrieval: an Introductory Manual to Principles and Practice*. 4th edition.- London: Library Association, 1992.- 310p.
4. Dialog Pocket Guide. *Information Solution Online*.
5. Gosling, Mary. *Learn reference work*. - Canberra: DocMatrPty Ltd, 1999. - 154p.
6. Grogan, Chris. *Practical reference work*. - London, Library Association, 1992.- 161p.
7. Harter S. P. *Online Information Retrieval*. - London : Academic Press, 1985.
8. Hye-soon Jeong. *Introductory to Data Bank*. Technical Cooperation Program. - Seoul: KINITI, 1992. p. 205-217.
9. Informatica.- M.: *Kniga*, 1986.- 301 tr. (tiếng Nga).
10. Kats William. *A introduction to Reference work*. Vol. I: Basic Information Sources. 7th ed.- NY.: McGraw-Hill, Inc, 1997. - 398 p.
11. Kats William. *A introduction to Reference work*. Vol. II: Reference services and reference processes. 7th ed. - NY.: McGraw-Hill, Inc, 1997. - 259p.
12. Knight Ridder Information Inc.: *Complete Databae Catalogue* . - 1997.

13. Kowalski, Gerald. *Information Retrieval System: Theory and Implementation*. - Boston, Dordrech. - London : Kluwer Academic Publishers, 1997. 282p.
14. Krol. *The Whole Internet Guide*. - New York: 1994.
15. Lankaster F.W. *Sistemu informasionnogo noisca*. - Minsk, 1972.- 308 tr. (tiếng Nga)
16. Meadow, Charles T. *Text Information Retrieval Systems*. - Toronto, Academic Press, Inc, 1992.-302p.
17. Mikhailov A.I., A.I. Chernui, Gliarevski S. *Osnovu informatiki* - M., Nauka, 1969 - 306tr. (Tiếng Nga).
18. Stair, R.M. *Principles of Information Systems: A Managerial Approach*.
19. Boyd & Fraser Publishing Company. *2d Edittion*.- London - New York: 1996.- 656 p.
20. STN - Guide of commands . *Scientific and Technical Network*.

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

16 Hàng Chuối - Hai Bà Trưng - Hà Nội

Điện thoại: (04) 9714896; (04) 9724770. Fax: (04) 9714899

Chịu trách nhiệm xuất bản:

Giám đốc: PHÙNG QUỐC BẢO

Tổng biên tập: NGUYỄN BÁ THÀNH

Biên tập: LÊ THUY

Chế bản: LÊ THUY

Trình bày bìa: QUANG HÙNG

TRA CỨU THÔNG TIN TRONG HOẠT ĐỘNG THƯ VIỆN - THÔNG TIN

Mã số: 2L-271 ĐH2008

In 1.000 cuốn, khổ 14.5 x 20.5 cm tại Công ty In & Văn hoá phẩm

Số xuất bản: 904 - 2008/CXB/04 - 158/ĐHQGHN, ngày 25/9/2008

Quyết định xuất bản số: 271 LK-XH/XB

In xong và nộp lưu chiểu quý IV năm 2008.



Giá: 35.000đ