

SO SÁNH CÁC NGÔN NGỮ TÌM TIN VỚI NGÔN NGỮ ĐỀ MỤC CHỦ ĐỀ

ThS. Lê Thị Ngọc Thu, Huỳnh Công Khanh
Thư viện khoa học tổng hợp TP. HCM

Ngôn ngữ tìm tin được sáng tạo ra để sử dụng trong hệ thống thông tin tư liệu, để mô tả nội dung tài liệu và để tìm thông tin. Có nhiều loại ngôn ngữ tìm tin (ngôn ngữ tư liệu) và có thể dựa vào nhiều dấu hiệu khác nhau để phân loại ngôn ngữ tư liệu, ngôn ngữ tìm tin.

***Dựa vào phạm vi bao quát đề tài của ngôn ngữ tư liệu, có:**

- *Ngôn ngữ tư liệu bách khoa, tổng hợp*: phản ánh đầy đủ các lĩnh vực tri thức, các ngành khoa học. Ví dụ: Bảng phân loại DDC, Bảng tra chủ đề của Thư viện Quốc hội Hoa Kỳ, Bộ từ khóa của Thư viện Quốc gia Việt Nam,...

- *Ngôn ngữ tư liệu đa ngành*: phản ánh một số ngành khoa học, một số lĩnh vực tri thức. Ví dụ: Bộ từ khóa của Trung Tâm Thông Tin Tư Liệu KH & CN Quốc gia Hà Nội,...

***Dựa vào nguyên tắc xây dựng ngôn ngữ tư liệu, có:**

- *Ngôn ngữ tư liệu kết hợp trước* (ngôn ngữ tư liệu tiền kết hợp). Đó là ngôn ngữ tư liệu mà những khái niệm đã được kết hợp trước khi cán bộ thư viện – thông tin hay người dùng tin sử dụng. Ví dụ: Bảng phân loại, Bảng đề mục chủ đề,...

- *Ngôn ngữ tư liệu kết hợp sau* (ngôn ngữ tư liệu hậu kết hợp). Đó là ngôn ngữ tư liệu mà những khái niệm được kết hợp ngay trong quá trình xử lý tài liệu của cán bộ thư viện hay trong quá trình tìm thông tin của người dùng tin. Ví dụ: Bộ từ khóa, Từ điển từ chuẩn (Thesaurus),...

***Dựa vào dấu hiệu tìm tin, có:**

- Ngôn ngữ tìm tin theo phân loại.
- Ngôn ngữ tìm tin theo từ khóa, từ chuẩn.
- Ngôn ngữ tìm tin theo chủ đề.

1. Ngôn ngữ tìm tin theo phân loại:

Ngôn ngữ phân loại là một hệ thống các ký hiệu tượng trưng được sử dụng để diễn đạt nội dung khái quát của tài liệu theo kết cấu của bảng phân loại, trong từng môn loại lại có sự phân chia chi tiết theo nguyên tắc từ chung đến riêng thể hiện mối quan hệ thứ bậc.

Ngôn ngữ phân loại được sử dụng từ lâu và rộng rãi trong công tác thư viện, tư liệu, lưu trữ và thông tin. Ngôn ngữ phân loại phổ biến hơn cả là DDC, UDC, BBK,...

Ngôn ngữ tìm tin theo phân loại có ưu và nhược điểm sau:

Ưu điểm:

- Logic chặt chẽ, hệ thống hóa cao trên cơ sở phân loại vốn tài liệu theo lĩnh vực tri thức.

- Phản ánh nội dung tài liệu chính xác.

- Tìm tin nhanh chóng, dễ dàng.

- Vai trò mục lục phân loại trong tìm tin theo truyền thống rất lớn vì nó đáp ứng nhu cầu tìm tin theo chuyên ngành. Người sử dụng thư viện có thể tìm tài liệu về các lĩnh vực khoa học trong mục lục phân loại. Việc ứng dụng tin học vào hoạt động thư viện – thông tin đã làm tăng khả năng tìm kiếm thông tin, truy cập thông tin theo phân loại tuy ít hơn nhưng không thể bỏ qua. Khi tìm thông tin ở những mảng đề tài lớn thì chọn tìm tin theo phân loại sẽ cho kết quả bao quát và đầy đủ hơn.

Sự kết hợp tìm thông tin giữa ký hiệu phân loại và từ khóa, từ chuẩn thường đạt kết quả tối đa.

Ví dụ: tìm toàn bộ tài liệu về “chăn nuôi đại cương” trong khung DDC có ký hiệu 636 tìm theo ký hiệu này thu được kết quả tối ưu những tài liệu viết về “chăn nuôi đại cương” bao gồm mọi khía cạnh. Nếu tìm theo từ khóa, từ chuẩn thì kết quả có thể rộng hơn hoặc hẹp hơn. Kết quả tìm tin rộng hơn khi từ khóa “chăn nuôi” bao gồm cả chăn nuôi đại cương và chăn nuôi chuyên ngành. Hoặc có những tài liệu về chăn nuôi hay liên quan tới chăn nuôi nhưng lại không có từ

khóa “chăn nuôi” mà lại có từ khóa khác như ”nuôi lợn”, “nuôi gia súc”, “nuôi gia cầm”, vv... Chúng ta có thể sử dụng đồng thời ngôn ngữ phân loại và ngôn ngữ từ khóa để kết hợp trong biểu thức tìm tin, điều này làm cho khả năng tìm tin phong phú hơn.

Nhược điểm:

- Tài liệu viết về cùng một vấn đề sẽ phản ánh ở nhiều môn loại khác nhau trong khung phân loại, điều này dẫn đến việc phân tán thông tin.

Ví dụ: theo khung phân loại DDC các tài liệu có nội dung về hóa học được phản ánh ở các ký hiệu trong khung như sau:

540	Hóa học
547	Hóa học hữu cơ
546	Hóa học vô cơ
660	Hóa học ứng dụng

Do vậy phải có sự hỗ trợ của bảng tra chủ đề chữ cái thì bảng phân loại mới đáp ứng được việc tiếp cận khác nhau với nội dung tài liệu.

- Nguyên tắc phân loại theo mối quan hệ cấp bậc và trật tự trong từng môn ngành tri thức chặt chẽ nên ngôn ngữ tìm tin theo phân loại khó phản ánh được các vấn đề mới, các lĩnh vực mới.

- Cấu trúc theo sơ đồ hình cây, chỉ thể hiện mối quan hệ cấp bậc trong từng môn ngành tri thức mà không phản ánh được mối quan hệ liên ngành, đa ngành giữa các vấn đề.

- Thực tế người sử dụng thư viện ít quan tâm và ít nhớ đến các ký hiệu phân loại.

2. Ngôn ngữ tìm tin theo từ khóa, từ chuẩn:

* Từ khóa là một trong các loại ngôn ngữ tư liệu hậu kết hợp. Các từ và tổ hợp từ phản ánh mẫu tìm của tài liệu mặc dù không phụ thuộc lẫn nhau nhưng khi cần thiết lại kết hợp với nhau.

Từ khóa được dùng để phản ánh nội dung của tài liệu đưa vào cơ sở dữ liệu và được dùng như là một dấu hiệu tìm tin theo nội dung. Từ khóa còn được dùng để xử lý và phản ánh nội dung của yêu cầu thông tin.

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

Ngôn ngữ từ khóa có những ưu nhược điểm sau:

Ưu điểm:

- Khả năng tập hợp tài liệu theo đối tượng nghiên cứu, không phụ thuộc vào lĩnh vực hoạt động.

- Có khả năng bổ sung từ mới trong quá trình mô tả. Cập nhật dễ và nhanh, phản ánh kịp thời sự vật, hiện tượng mới.

- Không đòi hỏi khả năng khái quát hóa cao nội dung tài liệu (so với chủ đề).

- Mô tả chi tiết nội dung tài liệu hơn (so với chủ đề).

Nhược điểm:

- Các vấn đề cùng một lĩnh vực bị tản mạn.

- Các từ khóa độc lập với nhau nên không mở rộng khả năng tìm kiếm thông tin của người dùng tin. Sự kết hợp các từ khóa trong quá trình tìm tin đôi khi dẫn đến kết hợp sai. Ví dụ: Thư viện trường học và trường học thư viện, khi tìm đều ra những thông tin giống nhau.

- Độ nhiễu tin lớn hơn (so với chủ đề).

* Thesaurus 1. Bảng từ mô tả. Một bảng liệt kê những thuật ngữ bao gồm những chữ đồng nghĩa và cho thấy được đẳng cấp cũng như mối liên hệ phụ thuộc giữa các từ, mà chức năng là cung cấp một ngữ vựng tiêu chuẩn, có kiểm soát sử dụng trong việc lưu trữ và truy dụng thông tin. Các thành phần của nó gồm một ngữ vựng chính và một ngữ vựng hướng dẫn. 2. Từ điển đồng nghĩa/ phản nghĩa. Một từ điển loại từ vựng, được sắp xếp theo thứ tự phân loại thường bao gồm những chữ đồng nghĩa và phản nghĩa. Thesaurus gần giống như bảng đề mục chủ đề. Đó là danh mục từ được xếp theo thứ tự chữ cái. Sau mỗi từ có thể là một số chỉ dẫn về các từ đồng nghĩa (synonyms), từ nghĩa rộng (Broad terms), từ nghĩa hẹp (Narrow terms), từ liên quan (Relative terms). Những từ trong thesaurus là những từ được chuẩn hoá và được quy định dùng để biểu thị một khái niệm nhất định.

Trên thế giới có một số từ điển chuẩn như: Thesaurus of psychological terms, ERIC Thesaurus,... Ở Việt Nam chưa có bộ từ điển từ chuẩn bằng tiếng Việt. Ưu nhược điểm của ngôn ngữ tìm tin theo từ chuẩn:

Ưu điểm:

- Là ngôn ngữ kết hợp sau nên rất linh hoạt, tiếp thu được những đặc điểm của các loại ngôn ngữ có kiểm soát, chẳng hạn của các khung phân loại, đặc biệt là hệ thống phân loại theo diện, trong khi vẫn phát triển các đặc điểm mới của riêng mình như hình thức trình bày đồ họa (hình cây, hình mũi tên,...), đồng thời cũng là ngôn ngữ phản ánh chính xác nội dung tài liệu.

- Có cấu trúc ngữ nghĩa, lựa chọn thuật ngữ ưu tiên trong số các từ đồng nghĩa và giới hạn ý nghĩa của các thuật ngữ để gây lầm lẫn bằng cách sử dụng các bổ ngữ và ghi chú phạm vi. Ngôn ngữ tìm tin theo từ chuẩn mở rộng khả năng tìm kiếm thông tin trong hệ thống tìm tin.

Nhược điểm:

- Cấu trúc chặt chẽ của từ chuẩn sẽ khó khăn khi bổ sung thêm từ mới.
- Từ chuẩn thường chỉ phản ánh một ngành hoặc một liên ngành có giới hạn, vì vậy phạm vi sử dụng hẹp.

3. Ngôn ngữ tìm tin theo chủ đề:

Ngôn ngữ tìm tin theo chủ đề có những ưu nhược điểm sau:

Ưu điểm:

*So sánh với ngôn ngữ tìm tin theo phân loại:
- Ngôn ngữ tìm tin theo chủ đề có khả năng tập hợp tài liệu theo đối tượng nghiên cứu, vấn đề, không phụ thuộc vào lĩnh vực hoạt động.

Ví dụ: chủ đề về Anh ngữ, chúng ta có

Anh ngữ

+Danh từ

+Động từ

+Ngữ pháp

+Từ vựng

Ngôn ngữ chủ đề giúp cho người dùng tin truy cập và tìm tin về một vấn đề nhanh chóng, đầy đủ, chính xác.

- Là ngôn ngữ hỗ trợ cho ngôn ngữ tìm tin theo phân loại (Bảng tra chủ đề chữ cái của bảng phân loại, ô tra chủ đề chữ cái của Mục lục phân loại).

*So sánh với từ khóa, từ chuẩn:

- Ngôn ngữ tìm tin theo chủ đề và ngôn ngữ tìm tin theo từ khóa đều được xây dựng trên cơ sở ngôn ngữ tự nhiên, để phản ánh và để tìm thông tin.

- Ngôn ngữ tìm tin theo chủ đề là ngôn ngữ tư liệu kết hợp trước còn ngôn ngữ tìm tin theo từ khóa là ngôn ngữ tư liệu kết hợp sau.

- Sự khác nhau cơ bản giữa hai ngôn ngữ này là ở dạng thức trình bày và nguyên tắc xử lý:

+Xử lý theo chủ đề: tức là xác định chủ đề chính – phụ đề. Phụ đề gồm 4 loại là phụ đề nội dung, phụ đề địa lý, phụ đề thời gian, phụ đề hình thức. Phụ đề dùng để phản ánh các khía cạnh khác nhau của chủ đề chính. Số lượng các đề mục chủ đề hạn chế so với số lượng các từ khóa.

+Xử lý theo từ khóa, từ chuẩn: là xây dựng một tập hợp các từ khóa, từ chuẩn. Những từ này phản ánh nội dung tài liệu và được chọn lọc từ trong chính văn hoặc nhan đề tài liệu. Số lượng từ khóa không hạn chế.

Ví dụ: tài liệu nói về “Kinh tế nông nghiệp của Việt Nam”.

Xử lý theo chủ đề:

Nông nghiệp – Khía cạnh kinh tế – Việt Nam

Xử lý theo từ khóa:

Nông nghiệp

Kinh tế

Việt Nam

- Tìm tin theo từ khóa phải xác định được các dấu hiệu cơ bản cần tìm kiếm, chọn các toán tử tìm và lập ra công thức tìm cho thích hợp để tìm đúng và đầy đủ tài liệu mình cần. Muốn vậy phải cần có sự hỗ trợ của máy tính, phương tiện kiểm soát từ vựng, người tìm tin phải có những kỹ năng nhất định về máy tính và kỹ năng về phương pháp tìm tin thì mới đảm bảo cho hiệu quả tìm tin.

Vì vậy, tìm tin theo từ khóa gắn liền với việc sử dụng máy tính điện tử. Trong khi tìm tin theo chủ đề có thể tìm tin ở các thư viện – cơ quan thông tin còn sử dụng các phương tiện tra cứu truyền thống song song với các phương tiện hiện đại. Ngôn ngữ tìm tin theo chủ đề còn có thể làm cơ sở để xây dựng và phát triển ngôn ngữ tìm tin theo từ khóa. Hiện nay ở các nước như Hoa Kỳ, Anh, Pháp,... ngôn ngữ tìm tin theo chủ đề ngày càng phát triển như Bảng đề mục chủ đề của thư viện Quốc hội Hoa Kỳ, RAMEAU thường xuyên được cập nhật khái niệm mới và được tiếp tục tái bản.

Sau đây là bảng so sánh những đặc điểm của đề mục chủ đề và từ khóa

TỪ KHÓA	ĐỀ MỤC CHỦ ĐỀ
Sử dụng bất kỳ thuật ngữ để mô tả nội dung đề tài.	Sử dụng từ vựng kiểm soát tiền kết hợp để mô tả nội dung của tài liệu (sách, bài báo) trong cơ sở dữ liệu.
Linh hoạt hơn trong tìm kiếm vì có thể kết hợp với nhau theo nhiều cách.	Kém linh hoạt hơn trong tìm kiếm vì cần biết chính xác thuật ngữ từ vựng kiểm soát.
Cơ sở dữ liệu tìm theo từ khóa ở bất kỳ nơi nào trong biểu ghi – không cần thiết kết nối với nhau.	Cơ sở dữ liệu tìm theo chủ đề chỉ trong đề mục chủ đề hay trong trường mô tả, nơi mà các từ có liên quan đã có.
Có thể kết quả quá nhiều hoặc quá ít kết quả.	Nếu quá nhiều kết quả – sử dụng phụ đề tập trung vào một khía cạnh của chủ đề rộng.

Nhược điểm:

- Ngôn ngữ tìm tin theo chủ đề ít có khả năng phản ánh tài liệu một cách có hệ thống mối quan hệ cấp bậc và trật tự trong từng môn ngành tri thức. Sắp xếp theo thứ tự chữ cái, bảng đề mục chủ đề chỉ có khả năng giới thiệu nội dung vốn tài liệu theo từng vấn đề cụ thể và những gì xoay quanh vấn đề đó.

- Các vấn đề cùng một lĩnh vực bị tản mạn.

Ví dụ: không thể tìm tất cả tài liệu về nông nghiệp trong ngành nông nghiệp.

- Đòi hỏi trình độ khái quát vấn đề cao (so với từ khóa).
- Mức độ chính xác bị hạn chế trong trường hợp đề mục chủ đề tổng quát

Tóm lại:

Mỗi ngôn ngữ tìm tin đều có ưu điểm và nhược điểm riêng. Ngôn ngữ tìm tin theo chủ đề và phân loại sử dụng được ở hệ thống tra cứu truyền thống và tra cứu tìm tin hiện đại. Trong khi ngôn ngữ tìm tin theo từ khóa, từ chuẩn chỉ áp dụng với tìm tin trên máy tính. Tìm tin theo chủ đề kém linh hoạt hơn so với tìm tin theo từ khóa, nhưng khả năng tập trung theo vấn đề trong mục lục chặt chẽ hơn so với từ khóa. Các phụ đề phản ánh góc độ của vấn đề, vì vậy có thể đáp ứng nhu cầu tra tìm tài liệu chuyên sâu.

Ngôn ngữ tìm tin theo chủ đề không hệ thống hoá cao như ngôn ngữ tìm tin theo phân loại nhưng tiện lợi cho người sử dụng. Ngôn ngữ tìm tin theo phân loại chỉ thể hiện sự tối ưu nếu có sự hỗ trợ của ô tra cứu chủ đề. Trong công tác phân loại, kỹ năng định chủ đề tốt sẽ giúp cho việc xác định ký hiệu phân loại thuận lợi hơn. Chính vì vậy, trong các thư viện và trung tâm thông tin cần phải sử dụng nhiều loại ngôn ngữ tìm tin để có sự hỗ trợ lẫn nhau giữa các loại ngôn ngữ tìm tin, đáp ứng nhu cầu tìm thông tin phong phú, đa dạng của người dùng tin. Do đó, ngôn ngữ tìm tin theo chủ đề cũng không thể thiếu trong các thư viện và trung tâm thông tin.