

TÌM HIỂU CẤU TRÚC CỦA FORMAT MARC

(Bài đăng trong Tập san Thư viện, số 2 – 2002)

PGS.TS.ĐOÀN PHAN TÂN
TRƯỜNG ĐẠI HỌC VĂN HOÁ HÀ NỘI

Quá trình ứng dụng tin học trong việc lưu trữ và khai thác tài liệu gắn liền với quá trình thực hiện tự động hoá công tác biên mục, một khâu quan trọng trong dây chuyền xử lý thông tin.

Thực chất của biên mục tự động là sử dụng một phần mềm tư liệu hoặc sử dụng phân hệ biên mục của phần mềm quản trị thư viện để tạo lập các biểu ghi cho một CSDL thư mục và tạo ra các mục lục thích hợp.

Trong biên mục tự động việc tạo lập biểu ghi thường là xử lý tiền máy và nhập dữ liệu, do con người thực hiện. Người ta nhập dữ liệu qua các khổ mẫu hiển thị trên màn hình theo kiểu xử lý văn bản. Còn việc tổ chức và sắp xếp biểu ghi và biên soạn các mục lục thì do máy tính thực hiện. Máy tính có thể in các phiếu mục lục và chế bản cho các ấn phẩm thư mục. Đó là sản phẩm đầu ra của CSDL thư mục.

Quy tắc mô tả thư mục theo tiêu chuẩn quốc tế ISBD chưa hoàn toàn thích nghi với cách xử lý tin học đối với các mô tả thư mục. Các phương pháp xử lý tin học đòi hỏi dữ liệu phải được cấu trúc hoá. Để máy tính có thể nhận biết được các dữ liệu thư mục, các chỉ dẫn không những phải tuân theo những quy tắc chặt chẽ, mà còn phải được trình bày một cách chính xác theo một khổ mẫu thống nhất. Ví dụ như kiểu ký tự sử dụng, độ dài của các trường dữ liệu, thứ tự sắp xếp v.v... Các quy tắc và cách trình bày đó tạo thành một format.

Format hay khổ mẫu là hình thức trình bày dữ liệu trên biểu ghi được cấu trúc hoá. Các dữ liệu trên biểu ghi được sắp xếp thành các trường, trường con, kết hợp với các mã số và các ký hiệu chỉ dẫn để điều hành sự sắp xếp này sao cho có thể nhận biết, trình bày và tìm kiếm dữ liệu bằng máy tính.

Năm 1966 lần đầu tiên Thư viện Quốc hội Mỹ xuất bản tập quy tắc của format MARC (Machine Readable Cataloging – Biên mục đọc máy). Format MARC là một mô tả có cấu trúc, dành riêng cho các dữ liệu thư mục được đưa vào máy tính điện tử.

Cấu trúc của format MARC là cấu trúc biểu ghi, trong đó các dữ liệu thư mục được sắp xếp trong các trường, có độ dài xác định, được mã hoá và trình bày theo một quy định chặt chẽ. Format MARC sử dụng các chữ số,

chữ cái, các ký hiệu ngắn gọn đặt ngay trong biểu ghi thư mục để đánh dấu và nhận biết các loại thông tin khác nhau trong mỗi biểu ghi.

Mỗi biểu ghi của format MARC bao gồm các trường (fields). Ngoài các trường dành cho các yếu tố mô tả thư mục theo AACR2 như : nhan đề, thông tin về trách nhiệm, thông tin về xuất bản, từng thư, đặc trưng số lượng, phụ chú, tóm tắt, v.v.... còn có các trường dành cho đề mục chủ đề, ký hiệu phân loại của Thư viện Quốc hội Mỹ, ký hiệu phân loại thập tiến Dewey,... Các trường này lại có thể chia nhỏ thành các trường con (subfields). Vì tên của trường thường khá dài, nên trong biểu ghi MARC mỗi trường được biểu diễn bằng một nhãn gồm 3 chữ số.

Để tiện trình bày, người ta tập hợp các nhãn trường thành từng nhóm. Các trường có nhãn bắt đầu bằng số “1” thuộc nhóm trường “1xx”, đó là các trường tiêu đề chính (tên cá nhân, tên tập thể, ...), các trường có nhãn bắt đầu bằng số “2” thuộc nhóm trường “2xx”, đó là các trường mô tả (nhan đề và thông tin trách nhiệm, thông tin về xuất bản, lần xuất bản,...), v.v...

Các nhãn trường thường được sử dụng là:

- 003 Chỉ số xác định tài liệu
- 010 Số kiểm tra của Thư viện Quốc hội Mỹ (LCCN)
- 020 Chỉ số sách quốc tế (ISBN)
- 022 Chỉ số ấn phẩm định kỳ quốc tế (ISSN)
- 082 Chỉ số phân loại thập phân Dewey
- 100 Tiêu đề chính, tên cá nhân (tác giả)
- 245 Nhan đề và thông tin về trách nhiệm
- 250 Thông tin về lần xuất bản
- 260 Thông tin về xuất bản, phát hành
- 300 Mô tả vật lý (đặc trưng số lượng)
- 440 Thông tin về từng thư
- 520 Chú giải hay tóm tắt
- 650 Điểm truy nhập chủ đề, đề mục chủ đề
- 700 Điểm truy nhập bổ sung, tên cá nhân

Thí dụ:

100	1#	\$a	Pirsig, Robert M.
-----	----	-----	-------------------

Trong đó 100 là nhãn trường tác giả, xác định như một điểm truy nhập chính là tên cá nhân. Còn ký hiệu 1# là chỉ thị và \$a là mã trường con mà ta sẽ đề cập dưới đây.

Danh sách tất cả các nhãn trường được in trong 2 tập *MARC 21 Format for Bibliographic Data* do Thư viện Quốc hội Mỹ xuất bản. Qua thống kê người ta thấy rằng chỉ có khoảng 10% số nhãn được sử dụng thường xuyên trong các biểu ghi MARC, 90% còn lại ít khi được sử dụng.

Ngoài nhãn trường, trong biểu ghi MARC một số trường còn được xác định bằng các *chỉ thị (indicators)*. Chỉ thị được mã hoá bằng hai chữ số (từ 0 đến 9) đi theo sau nhãn trường. Có trường chỉ dùng chữ số thứ nhất hoặc thứ hai, có trường dùng cả hai. Ở vị trí chỉ thị không được dùng, người ta quy ước thay bằng dấu #. Khi không cần đến chỉ thị người ta dùng ký hiệu ##.

Ví dụ:

245 14 \$a The emperor's new clothes / \$c adapted from Hans Christian Andersen and illustrated by Janet Stevens
--

Trong ví dụ trên, 245 là nhãn trường nhan đề và thông tin về trách nhiệm, chỉ thị thứ nhất “1”, có nghĩa là trong phiếu mục lục phải in từ “TITLE” vào trước nhan đề. Còn chỉ thị thứ hai “4”, có nghĩa là bỏ qua 4 ký tự đầu của nhan đề (*T, h, e* và *dấu trống*). Khi đó nhan đề “*The emperor's new clothes*” sẽ được ghi ra là “*emperor's new clothes*”.

Trong biểu ghi MARC, trường con được nhận biết bởi *mã trường con (subfield codes)*, đó là một ký tự bằng chữ in thường, đặt sau *dấu ngăn cách \$ (delimiter)*.

Ví dụ:

300 ## \$a 675p. : \$b ill. ; \$c 24cm
--

Trong đó 300 là nhãn trường mô tả vật lý của cuốn sách, bao gồm các trường con \$a (Số trang), \$b (Thông tin minh hoạ), \$c (Khổ, cỡ). Ký hiệu ## có nghĩa là chỉ thị không được xác định.

Ví dụ:

260 ## \$a Newyork : \$b Chelsea House, \$c 1986
--

Trong đó 260 là nhãn trường xuất bản, bao gồm các trường con: \$a (Nơi xuất bản), \$b (Nhà xuất bản), \$c (Năm xuất bản).

Dưới đây là ví dụ về các trường trong khổ mẫu nhập tin của một biểu ghi MARC:

NHÃN TRƯỜNG	CHỈ THỊ	MÃ TRƯỜNG CON	DỮ LIỆU
100	1#	\$a	Amosky, Jim.
245	10	\$a	Racoons and ripe com/
		\$c	Jim Amosky.
250	##	\$a	1st ed.
260	##	\$a	New york :
		\$b	Lothrop, Lee & Shepard Books,
		\$c	c1987.
300	##	\$a	25 p :
		\$b	col, ill. ;
		\$c	26 cm
520	##	\$a	Hungry racoons feast at night in a field of ripe corn.
650	#1	\$a	Racoons
901	##	\$a	8009
903	##	\$a	\$15.00

Trong đó 901 là nhãn trường *mã số vạch*, còn 903 là nhãn trường *giá*, được dùng trong phạm vi quốc gia và cục bộ.

Các nhãn trường, chỉ thị và mã trường con là các dấu hiệu để nhận biết và điều khiển cách bố trí các trường và trường con, do các chương trình quản trị CSDL quy ước khi xây dựng. Khi một biểu ghi thư mục đã được đánh dấu một cách chính xác và được lưu trữ dưới dạng một tệp dữ liệu trên máy tính thì chương trình quản trị CSDL sẽ đánh dấu và tạo khuôn dạng cho các thông tin này để in ra thành một bản thư mục, một phiếu mục lục hay hiển thị trên màn hình. Các chương trình này còn cung cấp công cụ tìm kiếm các thông tin thoả mãn yêu cầu đặt ra, dựa trên các điểm truy nhập nằm trong các trường của biểu ghi MARC.

Ngoài các trường dữ liệu thư mục là phần chính của biểu ghi thư mục, mỗi biểu ghi MARC còn có các thành phần cố định sau, xuất hiện ở đầu mỗi biểu ghi:

- a- *Đầu biểu (leader)*. Đầu biểu gồm 24 ký tự đầu tiên của biểu ghi, ứng với 24 vị trí, mỗi vị trí này được gán cho một ý nghĩa xác định và thể hiện bằng một mã (chữ in thường, chữ số, hoặc khoảng trống), cho ta biết các thông tin về trạng thái và các thuộc tính của biểu ghi như: độ dài biểu ghi, loại hình tài liệu, cấp thư mục, mức độ mã hoá, quy tắc mô tả được sử dụng (ISBD, AACR), ... Nhiều thông tin trong đầu biểu ghi là để dành cho máy tính sử dụng để nhận dạng biểu ghi.

Ví dụ: Giả sử một biểu ghi MARC có đầu biểu:

01401cam 2200265 a 4500

thì ý nghĩa cơ bản của nó là:

Vị trí	Ký tự	Ý nghĩa ký tự của vị trí
0- 4	01041	Độ dài biểu ghi: 1041 ký tự
5	c	Trạng thái biểu ghi: đã sửa chữa
6	a	Dạng tài liệu: văn bản in
7	m	Cấp thư mục: sách chuyên khảo
.....		
18	a	Quy tắc mô tả được sử dụng: AACR2
.....		

- b- *Danh mục (Directory)*. Tiếp theo đầu biểu là một loạt các tiểu dẫn (bằng các chữ số) có độ dài xác định gọi là danh mục. Danh mục này cho biết các nhãn trường có trong biểu ghi, độ dài của trường và vị trí bắt đầu của trường trong biểu ghi. Danh mục được xây dựng bởi máy tính từ biểu ghi thư mục, dựa trên các thông tin đã nhập.

Ví dụ: Đoạn tiểu dẫn sau đây trong danh mục

... 245003600354250001200390260003700402 ...

có nghĩa là: trong biểu ghi có trường với nhãn 245, có độ dài là 36 ký tự và bắt đầu từ vị trí thứ 354. Tiếp theo là trường có nhãn 250, có độ dài 12 ký tự và bắt đầu ở vị trí thứ 390 (36+354=390). Tiếp theo nữa là trường có nhãn 260, có độ dài 12 ký tự, bắt đầu từ vị trí 402

(402=12+390). Như vậy đoạn tiêu dẫn trên trong danh mục có cấu trúc và ý nghĩa như sau:

Nhãn	Độ dài	Bắt đầu từ vị trí
.....		
245	0036	00354
250	0012	00390
260	0037	00402
.....		

Như vậy mỗi biểu ghi MARC phải bao gồm các thành phần cơ bản sau:

- Đầu biểu gồm 24 ký tự
- Một danh mục các trường dữ liệu mà với mỗi trường dữ liệu phải bao gồm một nhãn với 3 chữ số, độ dài của trường dữ liệu và vị trí của ký tự đầu tiên.
- Các dữ liệu tương ứng với các trường trong danh mục, chứa các dữ liệu thư mục cần xử lý với độ dài thay đổi.

Trong biểu ghi MARC, các trường cách nhau bởi dấu phân cách trường. Trong MARC21, dấu phân cách trường ký hiệu là ^ . Dấu phân cách trường đặt ở cuối mỗi trường. Người ta dùng ký hiệu \ để đánh dấu kết thúc mỗi biểu ghi.

Format MARC do Thư viện Quốc hội Mỹ xây dựng đầu tiên và sử dụng nên còn gọi là USMARC. Cấu trúc biểu ghi của format MARC tạo ra nhiều khả năng cho việc sắp xếp, chọn lọc, đánh chỉ số, tìm tin, hiệu đính và biên soạn, in ấn các ấn phẩm thư mục, các mục lục.

Sau khi được chỉnh lý vào năm 1968, format MARC là cơ sở cho sự ra đời của một loạt các format quốc gia, như: CAN.MARC của Canada, UK.MARC của Anh, AUS.MARC của Úc, INTERMARC của Pháp, IBERMARC của Tây Ban Nha. Mỗi hệ thống biên mục sử dụng format MARC đều có hướng dẫn riêng về hệ thống phân cách và hình thức trình bày các biểu ghi của mình.

Năm 1997 USMARC của Thư viện Quốc hội Mỹ kết hợp với CANMARC của Thư viện Quốc gia Canada tạo thành MARC21 và trở thành format chuẩn được nhiều phần mềm quản trị thư viện sử dụng.

* * *

TÀI LIỆU THAM KHẢO

1- Understanding MARC Bibliographic: Machine Readable Cataloguing/ published by the Cataloguing Distribution Service, Library of Congress, in collaboration with The Follett Software Co. in Betty Furrie, 2000.- 51p.

2- Manuel UNIMARC: version française/ IFLA, Ed. Saur, 1991.- 414p.