

# **PHƯƠNG PHÁP PHÂN LOẠI VÀ MÃ HÓA THÔNG TIN KINH TẾ XÃ HỘI PHỤC VỤ QUẢN LÝ**

**Phạm Hồng Vân - Viện KHTK**

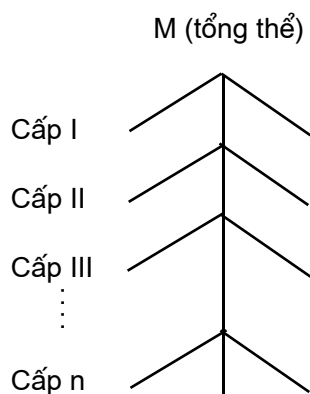
**P**hân loại và mã hóa là một khâu quan trọng của công tác đảm bảo thông tin trong quá trình hoàn thiện các hệ thống thông tin quản lý. Vì vậy, có thể coi hệ

thống phân loại và mã hóa là một bộ phận cấu thành của hệ thống thông tin nói chung, có chức năng chủ yếu là phục vụ quản lý. Trong phạm vi bài này chỉ đề cập

đến một số phương pháp phân loại và mã hóa cơ bản ứng dụng để xây dựng các hệ thống danh mục phục vụ công tác thống kê.

**1. Phân loại:** Là sự sắp xếp có trật tự các đối tượng của một tập hợp theo một tiêu thức được chọn. Tiêu thức này là đặc tính hoặc thuộc tính của đối tượng cần phân loại nó cho phép phân biệt các đối tượng cần nghiên cứu với nhau. Có hai phương pháp phân loại:

**a. Phân loại thứ bậc:** Phân loại thứ bậc là phân chia một tập liên kết phụ thuộc lẫn nhau một cách có cấp bậc. Việc phân chia này sẽ tiếp tục mãi mãi và mỗi lần phân chia dựa vào những tiêu thức cụ thể, mỗi nhóm, mỗi cấp có thể sử dụng vào những tiêu thức cụ thể, mỗi nhóm, mỗi cấp có thể sử dụng các tiêu thức khác nhau.



Đặc điểm của phương pháp phân loại này:

- Cho phép phân tổng thể nghiên cứu thành nhiều nhóm, nhiều cấp tùy theo yêu cầu quản lý và phân tích tổng thể. Số cấp biểu hiện mức độ chi tiết của tổng thể; càng nhiều cấp mức độ chi tiết càng sâu.

- Hệ thống phân loại này có tính logic chặt chẽ.

- Có thể xử lý các thông tin bằng phương pháp thủ công hoặc bằng phương tiện kỹ thuật hiện đại một cách thuận lợi. Tuy nhiên, phương pháp phân loại này kém tính mềm dẻo và uyển chuyển, thường phù hợp với việc phân loại các tổng thể tương đối ổn định. Nhưng trong thực tế các hiện tượng kinh tế xã hội luôn luôn biến động theo thời gian. Quá trình phát triển dẫn tới có những phần tử mất đi, ngược lại có những phần tử mới xuất hiện. Cho nên khi phân loại người ta thường để mã dự trữ cho các phần tử mới phát sinh mà không làm phá vỡ kết cấu chung của bảng danh mục.

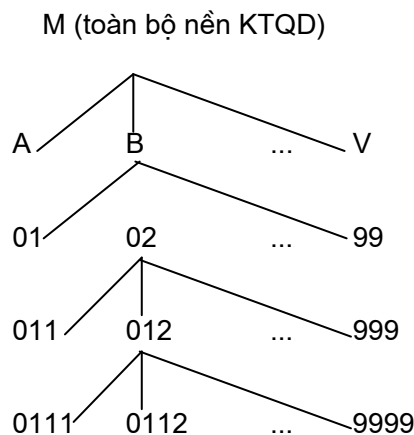
Hệ thống phân ngành kinh tế quốc dân (KTQD) ban hành theo Quyết định số 143/TCTK/PPCĐ, ngày 22 tháng 12 năm 1993 đã sử dụng phương pháp phân loại thứ bậc. Toàn bộ nền KTQD là một tập hợp được chia thành 4 cấp:

Cấp I: Từ mã A đến V

Cấp II: Từ mã 01 đến 99

Cấp III: Từ mã 011 đến 999

Cấp IV: Từ mã 0111 đến 9999



Ở mỗi cấp khác nhau có thể chọn các tiêu thức đặc trưng khác nhau. Ví dụ ngành cấp I được phân chia thành 20 ngành căn cứ vào tiêu thức tổng hợp nhất đó là các đặc trưng của quá trình hoạt động sản xuất kinh doanh, bao gồm:

- Đặc trưng của hàng hóa và dịch vụ được sản xuất ra
- Công dụng của sản phẩm hàng hóa
- Đầu vào (nguồn nguyên liệu) và qui trình công nghệ sản xuất. Ngành cấp I, cấp III và cấp IV cũng được sử dụng có chọn lọc 1 trong số các đặc trưng của quá trình hoạt động sản xuất.

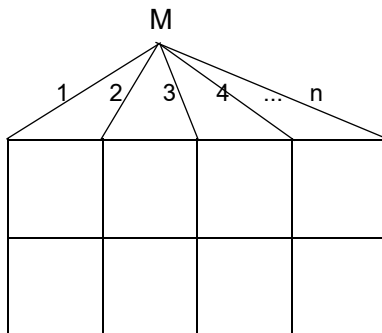
**b. Phân loại fasét:** Phân loại fasét là phân chia một tập các đối tượng của tổng thể chung theo các fasét, mỗi fasét đặc trưng cho một thuộc tính của tổng thể. Khác với phân loại thứ bậc, phân loại fasét có tính mềm dẻo và linh hoạt do vậy khi thêm bớt một fasét mới hoặc thay đổi trật tự của các fasét không làm ảnh hưởng đến cấu trúc chung của hệ thống phân loại. Điều này cho phép ta có thể chọn fasét sao cho đáp ứng được yêu cầu quản lý và sử dụng. Phân loại fasét có thể minh họa như sau:

F: fasét

F<sub>1</sub>: Có các nhóm 1, 2, 3, 4...n

F<sub>2</sub>: Có các nhóm 1, 2, 4

F<sub>3</sub>: Có các nhóm 2, 3, 4



Hệ thống danh mục xí nghiệp cơ quan sử dụng phương pháp phân loại fasét. Các đối tượng phân loại ở đây là các xí nghiệp, cơ quan còn thuộc tính của đối tượng là cấp quản lý, ngành kinh tế, thành phần kinh tế, qui mô v.v...

Khi nói đến hệ thống danh mục ngoài thành phần phân loại, sắp xếp các thông tin theo mục đích sử dụng cần phải quan tâm đến việc xây dựng hệ thống mã hóa sao cho phù hợp nhằm tạo điều kiện thuận lợi cho công tác xử lý, truyền đưa và cung cấp thông tin bằng các phương tiện hiện đại.

**2. Mã hóa:** Là quá trình gán cho mỗi đối tượng của tổng thể đã cho một bộ các ký hiệu theo các qui tắc xác định nhằm phục vụ cho việc xử lý, truyền đưa thông tin trong điều khiển nền KTQD. Bộ các ký hiệu có thể bằng tập chữ cái, chữ số, tín hiệu điện, tín hiệu âm thanh, tín hiệu cơ v.v... Cơ sở của mã hóa phụ thuộc vào số lượng đối tượng cần được mã hóa, nó có thể là các chữ trong bảng chữ cái (đối với mã chữ), cơ số thập phân, bách phân hoặc thiên phân v.v...(đối với mã số). Các ký hiệu có thể dùng để mã hóa cho đối tượng hoặc cho tính chất. Trong trường hợp nghiên cứu toàn bộ các ngành, các doanh nghiệp trong nền KTQD thì ngành và doanh nghiệp là đối tượng phân loại, còn tính chất là các thuộc tính chứa trong các đối tượng này. Có hai phương pháp mã hóa:

**a. Phương pháp mã mô tả** (của đăng ký): Là gán cho mỗi tính chất một số thứ tự mà không cần có sự phân loại, sắp xếp theo một tiêu thức nhất định. Ưu điểm của mã này là có dung tích lớn nhất vì nó có mối liên hệ hàm mũ.

Gọi: a là cơ số mã  
r số ngăn

M dung tích

Lúc đó với mã có r ngăn bằng nhau dung tích mũ sẽ là  $M_d = a^r$  và đối với dung

tích mã không đều  $M_d = a \left( \frac{a^r - 1}{a - 1} \right)$

**b. Phương pháp mã hóa song song:** Là loại mã hóa dựa trên cơ sở của phân loại fasét, mỗi ngăn dành cho phân loại một fasét.

Dung tích của mã song song:  $M = a.r$

Trong trường hợp mã song song theo khối:

$$M = \frac{r}{n} \cdot a^n$$

n: Mỗi khối

$a^n$ : Số tổ hợp của một khối

$\frac{r}{n}$ : Số khối

	Ngành	TPKT	Cấp QL	Lãnh thổ
XN				

Mã đăng ký

**c. Mã liên tiếp:** Mã hóa liên tiếp dựa trên cơ sở phân loại thứ bậc, mỗi cấp dùng một ngăn, các dấu hiệu của mỗi ngăn phụ thuộc vào dấu hiệu của các ngăn trên nó.

Dung tích của mã liên tiếp:

$$M = a + a^2 + \dots + a^r = a \frac{(a^r - 1)}{a - 1}$$

- Nếu mã liên tiếp mỗi khối đều l ngăn ta có:

$$M_l = a \frac{(a^{\frac{r}{l}} - 1)}{(a^l - 1)}$$

Mỗi khối phản ánh  $\frac{r}{l}$  tính chất

- Nếu mã hóa liên tiếp theo khối không đều ta có:

$$M_l' = a^{l_1} + a^{l_1+l_2} + \dots + a^{l_1+l_2+\dots+l_i}$$

**d. Phương pháp kết hợp giữa mã song song theo khối m ngăn và mã liên tiếp theo khối n ngăn:**

$$\text{Dung tích } M_m = \frac{r}{m} \cdot \frac{a(a^m - 1)}{a - 1}$$

Nếu: m = 1 trở về mã song song theo ngăn

m = r trở về mã liên tiếp

**Mã xí nghiệp cơ quan:**

Nếu dùng các khối không đều có độ dài mã từ 1 ÷ m

$$M'_m = \sum_{i=1}^m k_i \frac{a(a^i - 1)}{a - 1}$$

Khi phân tích đánh giá ưu nhược điểm của từng phương pháp phân loại chúng ta có thể phân tích riêng từng loại phương pháp, còn chọn phương pháp mã hóa nào cho phù hợp là dựa vào tính chất của đối tượng cần mô tả và dung lượng của tổng thể cần phân loại. Trong thực tế người ta thường sử dụng phương pháp mã hóa kết hợp: Mã song song theo khối, trong khối mã liên tiếp; mã liên tiếp theo khối, trong khối mã song song; mã liên tiếp trong một số ngăn đầu và mã song song trong một số ngăn còn lại v.v...