

# **Проектирование классификаторов технико- экономической информации**

**Лекция 5. ч. 1. Тема 1**

# Учебные вопросы

- Основные понятия классификации экономической системы
- Понятия и основные системы кодирования экономической информации

# Литература

- Романов В.П., Емельянова Н.З., Партыка Т.Л. Проектирование экономических информационных систем: методология и современные технологии: Учебное пособие / В.П. Романов, Н.З. Емельянова, Т.Л. Партыка. – М.: «Экзамен», 2005. – с. 50-59

# Первый учебный вопрос

## *Основные понятия классификации экономической системы*

Основные особенности **экономической информации**:

- проектирование классификаторов технико-экономической информации;
- большие объемы (до нескольких сотен млн. символов в год для среднего предприятия);
- большая ее часть имеет символьное представление, слабо приспособленное для логической и арифметической обработки;
- высокий уровень стоимостных и трудовых затрат на ее поиск и обработку;

Чтобы приспособить **экономическую информацию** для эффективного поиска, обработки на компьютере и передачи по каналам связи, ее **необходимо представить в цифровом коде**.

С этой целью ее нужно:

- упорядочить (классифицировать),
- формализовать (закодировать) с использованием классификатора.

**Классификатор** — это документ, с помощью которого осуществляется **формализованное описание экономической информации** в ЭИС, содержащей:

- наименования объектов,
- наименования классификационных группировок,
- их кодовые обозначения.

Экономическая информация существует в двух формах:

- экономических показателей,
- документов.

## ***Экономический показатель***

является **составной единицей информации**, отражающей **количественную характеристику** некоторого **процесса предметной области**, т.е. **реквизитом-основанием** вместе с однозначно определяющими его качество **реквизитами-признаками**.

**Реквизиты-основания** подразделяются по типу алгоритмов их получения на **количественные, стоимостные, проценты, удельные веса** и др.

Множество **реквизитов-признаков** по степени формализации делятся на:

- **справочные** — как правило, наименования предназначены для понимания показателя пользователем-экономистом;
- **группировочные** — это закодированные аналоги справочных признаков, предназначенные для логической обработки информации на компьютере.



Основными объектами классификации и кодирования являются **справочные реквизиты-признаки**, описывающие:

- объекты,
- процессы,
- место,
- время выполнения процессов,
- субъекты и объекты действия, отражаемые в показателе.

К объектам классификации и кодирования относятся также:

- *наименования показателей и документов;*
- *наименования компонентов* проекта ЭИС, в т.ч. файлов, задач, подсистем, программных модулей и др.

**Целью разработки** классификаторов является **установление соответствия** между значениями справочных или описательных признаков какого-либо элемента или процесса и значениями группировочных признаков,

Например, между значениями реквизитов «Фамилия И.О. рабочего» и «Табельный номер» рабочего или «Наименование материала» и «Код материала».

Для кодирования объектов необходимо их упорядочить по некоторым признакам.

Результат упорядоченного распределения объектов заданного множества носит название **классификации**, а совокупность правил разбиения множества на подмножества называется **системой классификации**.

Распределение объектов классификации в соответствии с принятой системой классификации носит название **процесса классифицирования**.

Свойство или характеристика объекта классификации, которое позволяет установить его сходство или различие с другими объектами классификации, называется **признаком классификации**.

Множество или подмножество, объединяющее часть объектов классификации по одному или нескольким признакам, носит название ***классификационной группировки***.

***Основанием классификации*** называется **признак**, по которому ведется **разбиение множества на подмножества** на определенной ступени классификации.

***Ступень классификации*** — это результат очередного **распределения объектов одной классификационной группировки**, расположенных на одних и тех же ступенях итерационности.

***Глубина системы классификации*** — это количество уровней классификации, допустимое в данной системе.

Каждая система классификации характеризуется следующими свойствами:

- гибкостью системы;
- емкостью системы;
- степенью заполненности системы (коэффициент заполненности).

***Гибкость системы*** — это способность допускать включение новых признаков, объектов без разрушения структуры классификатора. Гибкость определяется временем жизни ( $T_{ж}$ ) системы.

***Емкость системы*** — это наибольшее количество классификационных группировок, допускаемое в данной системе классификации ( $P$ ).

**Степень заполненности системы ( $K_{зап}$ )** определяется как частное от деления фактического количества группировок ( $Q_{ф}$ ) на величину емкости системы ( $P$ ):

$$K_{зап} = Q_{ф} / P.$$



В настоящее время чаще всего применяются два типа систем классификации: **иерархическая** и **многоаспектная**.

Характерными особенностями ***иерархической системы*** являются:

- наличие в системе ограниченного количества признаков классификации;
- соподчиненность признаков классификации, что выражается разбиением каждой классификационной группировки, образованной по одному признаку, на их множество по нижестоящему (подчиненному) признаку.

При построении иерархической системы классификации:

- выделяется некоторое множество объектов, подлежащее классифицированию,  $M_\sigma$ , для которого определяются полное множество признаков классификации  $G$  и их соподчиненность друг другу,
- производится разбиение исходного множества объектов на классификационные группировки на каждой ступени классификации.

Необходимо соблюдать следующие ограничения:

- получающиеся на каждом уровне классификационные группировки должны составлять исходное множество объектов  $M_j$ ;
- классификационные группировки  $X_{jk}$  на каждой ступени не должны пересекаться;
- классификация на каждой ступени должна проводиться только по одному признаку (G).

## Положительные стороны:

- логичность,
- простота построения,
- удобство логической и арифметической обработки.

## Недостаток:

- жесткая структура классификации, не позволяющей вносить новые признаки или изменять их последовательность.

Гибкость этой системы обеспечивается только за счет ввода большой избыточности в ветвях

**Аспект** — точка зрения на объект классификации, который характеризуется одним или несколькими признаками.

**Многоаспектная система** — это система классификации, которая использует параллельно несколько независимых признаков (аспектов) в качестве основания классификации.

Существуют два типа многоаспектных систем: фасетная и дескрипторная.

**Фасет** — это аспект классификации, который используется для образования независимых классификационных группировок.

**Дескриптор** — ключевое слово, определяющее некоторое понятие, которое формирует описание объекта и обозначающее принадлежность этого объекта к классу, группе и т.д.

# Особенности построения *фасетной* системы :

- имеется некоторое множество классифицируемых объектов ( $M_o$ );
- это множество можно рассматривать в нескольких аспектах, каждый из которых может характеризоваться одним или несколькими признаками, образующими фасет  $\Phi_r$ ;

# Особенности построения *фасетной* системы :

- устанавливается некоторый порядок следования фасетов с помощью фасетной формулы (при этом последовательность фасетов определяется по частоте обращения к этим фасетам на некотором множестве заданных задач):

$$F = (\Phi_1, \dots, \Phi_2, \dots, \Phi_r, \dots, \Phi_R);$$

# Особенности построения *фасетной* системы :

- определяется количество подмножеств классификационных группировок, число которых определяется числом задач, обращающихся при своем решении к тем или иным фасетам.

Внутри фасета значения признаков могут просто перечисляться в некотором порядке или образовывать сложную иерархическую структуру, если существует соподчиненность выделенных признаков.



## Преимущества данной системы:

- большая емкость,
- высокая степень гибкости, поскольку при необходимости можно вводить дополнительные фасеты и изменять их место в формуле.

## Недостатки:

- сложность структуры,
- низкая степень заполненности.

Для поиска показателей и документов по набору содержательных признаков используется **информационный язык дескрипторного типа**, который характеризуется совокупностью:

- терминов,
- дескрипторов или лексикой,
- набором отношений между терминами.

В зависимости от того, на каком этапе фиксируются все возможные выражения, языки делятся на **предкоординированные** и **посткоординируемые**.

**Предкоординированными** называются языки, в которых на стадии разработки выделяются все высказывания в терминах этих языков и тем самым заранее определяются постоянные отношения между терминами.

Для **посткоординируемых** языков характерна предварительная фиксация лишь постоянных отношений. Все высказывания образуются при использовании лексики данного языка и его грамматики.

К языкам классификационного типа можно отнести разработанный ныне ***Общесистемный классификатор технико-экономических показателей (ОКТЭП).***

Он содержит:

- способ упорядоченного представления системы показателей,
- средства их взаимной увязки на основе многоаспектной классификации этих показателей.

Общие **недостатки** информационных языков классифицированного типа:

- слабая приспособленность к новым, заранее не предусмотренным условиям функционирования систем,
- возможность составления запросов на этих языках только регламентированного содержания.

Эти недостатки отсутствуют у языков посткоординированного типа, к которым относятся **дескрипторные языки**, основанные на применении метода координатного или ассоциативного индексирования.

***Дескриптор*** — это термин естественного языка (слово или словосочетание), используемый при описании документов или показателей.

Он имеет самостоятельный смысл и неделим без изменения своего значения.

## ***Второй учебный вопрос***

### ***Понятия и основные системы кодирования экономической информации***

Для полной формализации экономической информации проводят следующую процедуру — кодирование.

***Кодирование*** — это процесс присвоения условных обозначений объектам и классификационным группам по соответствующей системе кодирования.

***Система кодирования*** — это совокупность правил обозначения объектов и группировок с использованием кодов.

**Код** — это условное обозначение объектов или группировок в виде знака или группы знаков в соответствии с принятой системой.

Код базируется на определенном **алфавите** (некоторое множество знаков).

Число знаков этого множества называется **основанием** кода.

Различают следующие типы алфавитов:

*цифровой,*

*буквенный,*

*смешанный.*



Код характеризуется следующими параметрами:

- длиной ( $Z$ );
- основанием кодирования ( $A$ );
- структурой, под которой понимают распределение знаков по признакам и объектам классификации;
- степенью информативности ( $I$ ), рассчитываемой как частное от деления общего количества признаков ( $K$ ) на длину кода ( $L$ ):

$$I = R / L;$$

Код характеризуется :

- коэффициентом избыточности ( $K_{изб}$ ), который определяется как отношение максимального количества объектов ( $Q_{max}$ ) к фактическому количеству объектов ( $Q_{факт}$ )

$$K_{изб} = Q_{max} / Q_{факт}$$

Все системы кодирования можно сгруппировать в два подмножества:

- регистрационные,
- классификационные системы.

Особенностью **регистрационных систем кодирования** является их независимость от применяемых систем классификации.

Регистрационные коды состоят из двух частей: **информационной** и **контрольной**, предназначенной для защиты передаваемой информации от ошибок.

Контрольная часть может рассчитываться по различным алгоритмам, в частности наиболее употребляемыми являются следующие формулы их расчета:

$$K = M - \left[ \sum X_i / M \right],$$
$$K = M - \left[ \sum X_i \times V_i / M \right],$$

где  $M$  — модуль (простое число, делящееся на единицу и на само себя);

$X_i$  — информационные разряды,

$i$  — номер разряда;

$V$  — вес информационного разряда.

К регистрационным системам относятся **порядковая** и **серийная**.

**Порядковая** —наиболее простая - последовательное присвоение каждому объекту кодируемого множества  $M$  номера его порядка, т.е. присвоение цифр натурального ряда в порядке расположения объектов.

Этот порядок может быть случайным или определяться после предварительной группировки объектов, например по алфавиту.

Как правило, порядковую систему применяют для кодирования малозначных, устоявшихся и простых множеств объектов, не требующих предварительной

## ***Серийная (серийно-порядковая)***

система кодирования отличается от ***порядковой*** тем, что номенклатура кодируемых объектов  $M$  предварительно должна быть разбита на группировки по одному признаку и каждой группировке должна быть отведена серия кодовых обозначений, в пределах которой каждому элементу присваивается свой код по порядку.

***Классификационные коды*** используют для отражения классификационных взаимосвязей объектов и группировок и применяются в основном для сложной логической обработки экономической информации на компьютере

Группу классификационных систем кодирования можно разделить **на две подгруппы** в зависимости от того, какую систему классификации используют для упорядочения объектов

## ***Последовательные системы***

кодирования характеризуются тем, что они базируются на предварительной классификации по иерархической системе классификации, в результате использования которой коды нижестоящих группировок образуются путем добавления кодов к кодам вышестоящих группировок.



## ***Параллельные системы***

кодирования строятся на основе **фасетной системы классификации** и коды группировок по фасетам формируются независимо друг от друга.

Последовательные и параллельные системы кодирования строятся на базе разрядной или комбинированной системы кодирования

***Разрядная система*** применяется для кодирования объектов, определяемых несколькими соподчиненными признаками, используемыми для решения экономических задач.

Кодируемые объекты систематизируются по классификационным признакам на каждой ступени классификации, каждому признаку отводится определенное число разрядов, в пределах которых кодирование группировок начинается с единицы.

## ***Комбинированная система***

кодирования, применяется для кодирования больших номенклатур (перечней) объектов

Эта система базируется на сочетании принципов построения таких систем кодирования, как **разрядная, серийная, порядковая и кода повторения.**

**Код повторения (мнемокод)** — это буквенные или буквенно-цифровые коды, которые характеризуются тем, что в структуру кода переносят часть символьных обозначений объектов с целью повышения мнемоничности кода или для сокращения его длины.

Выбор конкретной системы кодирования зависит от:

- объема кодируемой номенклатуры, ее стабильности,
- задач, стоящих перед системой,
- показателей эффективности обработки информации при использовании какой-либо системы.