## Основные понятия классификации технико-экономической информации

Для организации работы с информацией её нужно сначала упорядочить (классифицировать), а затем формализовать (закодировать) с использованием классификатора.

Классификатор – это документ, с помощью которого осуществляется формализованное описание экономической информации в ЭИС, содержащий наименования объектов, наименования классификационных группировок и их кодовые обозначения.

Экономическая информация существует в двух формах: в форме экономических показателей и документов.

Экономический показатель является составной единицей информации, отражающей количественную характеристику некоторого процесса предметной области – реквизит-основание вместе с однозначно определяющими его качествами реквизитами-признаками:



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

Множество *реквизитов-признаков* по степени формализации делится на два подмножества:

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

Основными объектами классификации и кодирования являются справочные реквизиты-признаки, описывающие процессы, место, время выполнения процессов, субъекты и объекты действия, отражаемые в показателе.

Например, к числу наименований элементов можно отнести наименования материальных, трудовых, денежных, энергетических ресурсов, основных средств, готовой продукции и услуг. К числу наименований процессов относятся наименования функций управления, деловых процессов, операций поступления сырья и материалов, отпуска их в производство, производства и выпуска готовой продукции или оказания услуг, процессов выполнения заказов, обслуживания клиентов, хранения, реализации готовой продукции, расчетов с поставщиками и покупателями, получения оплаты за реализованную продукцию и т.д.

К объектам классификации и кодирования относятся также наименования показателей и документов. Помимо этого к объектам классификации и кодирования относят также наименования компонентов проекта ЭИС, в том числе файлов, задач, подсистем, программных модулей и др.

Целью разработки классификаторов является установление соответствия между значениями справочных или описательных признаков, например:

«Фамилия И.О. рабочего» ↔ «Табельный номер» рабочего

«Наименование материала» ↔ «Код материала».

Классификация – это разделение множества объектов на подмножества по их сходству или различию в соответствии с принятыми методами. Фиксирует закономерные связи между классами объектов. Система классификации позволяет сгруппировать объекты и выделить определенные классы, которые будут характеризоваться рядом общих свойств.

Т. о. система классификации - совокупность правил распределения объектов множества на подмножества.

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**.

**Классификатор** (справочник) — это документ, с помощью которого осуществляется формализованное описание информации в ИС, содержащей наименования объектов, наименования классификационных группировок и их кодовые обозначения.

По сфере действия выделяют следующие виды классификаторов:

1. Международные классификаторы входят в состав Системы международных экономических стандартов (СМЭС) и обязательны для передачи информации между организациями разных стран мирового сообщества;
2. Общегосударственные (общесистемные) классификаторы, обязательны для организации процессов передачи и обработки информации между экономическими системами государственного уровня внутри страны;
3. Отраслевые классификаторы используют для выполнения процедур обработки информации и передачи ее между организациями внутри отрасли;
4. Локальные классификаторы используют в пределах отдельных предприятий.

Каждая система классификации характеризуется следующими свойствами:

* гибкостью системы;
* емкостью системы;
* степенью заполненности системы.

**Гибкость системы** — это способность допускать включение новых признаков, объектов без разрушения структуры классификатора. Необходимая гибкость определяется временем жизни системы.

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

**Степень заполненности системы** Кзап определяется как частное от деления фактического количества группировок Qф на величину емкости системы:

Кзап=Qф/P.

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

### Иерархическая система классификации

Характерными особенностями иерархической системы являются:

* возможность использования неограниченного количества признаков классификации;
* соподчиненность признаков классификации, что выражается разбиением каждой классификационной группировки, образованной по одному признаку, на множество классификационных группировок по нижестоящему (подчиненному) признаку.

Правила построения иерархической системы классификации:

* каждая классификационная группировка должна делится только по одному основанию деления;
* получаемые в результате деления группировки должны исключать друг друга т.е. не повторятся;
* сумма классификационных группировок должна составлять первоначальное классифицируемое множество.



Таким образом, классификационные схемы, построенные на основе иерархического принципа, имеют неограниченную емкость, величина которой зависит от глубины классификации (числа ступеней деления) и количества объектов классификации, которое можно расположить на каждой ступени. Количество же объектов на каждой ступени классификации определяется основанием кода, то есть числом знаков в выбранном алфавите кода. (Например, если алфавит – двузначные десятичные числа, то можно на одном уровне разместить 100 объектов). Выбор необходимой глубины классификации и структуры кода зависит от характера объектов классификации и характера задач, для решения которых предназначен классификатор.

Плюсы: логичность, простоту ее построения и удобство логической и арифметической обработки.

Минус - жесткость классификационной схемы - заранее установленный выбором признаков классификации и порядок их использования по ступеням классификации. Т.о. при изменении состава объектов классификации, их характеристик или характера решаемых требуется коренная переработка классификационной схемы.

Поэтому при разработке классификаторов следует учитывать, что иерархический метод классификации более предпочтителен для объектов с относительно стабильными признаками и для решения стабильного комплекса задач.



Классификатор минералов

### Многоаспектные системы классификации

Недостатки, отмеченные в иерархической системе, отсутствуют в других системах, которые относятся к классу многоаспектных систем классификации.

**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;
* дескрипторная.

#### Фасетная система классификации

**Фасет** — это аспект классификации, который используется для образования независимых классификационных группировок. **Дескриптор** — ключевое слово, определяющее некоторое понятие, которое формирует описание объекта и дает принадлежность этого объекта к классу, группе и т.д.

Под фасетным методом классификации понимается "параллельное разделение множества объектов на независимые классификационные группировки". При этом методе классификации заранее жесткой классификационной схемы и конечных группировок не создается. Разрабатывается лишь система таблиц признаков объектов классификации, называемых фасетами.

Особенности построения:

* множество классифицируемых объектов можно рассматривать в нескольких аспектах, каждый из которых может характеризоваться одним или несколькими признаками, образующими фасет *Фn*;
* 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=(Ф1,Ф2,...,Фr,...,ФR)*;
* определяется количество подмножеств классификационных группировок, число которых определяется числом задач, обращающихся при своем решении к тем или иным фасетам.



Для группировки осуществляется выборка необходимых признаков из фасетов и их объединение в определенной последовательности. Внутри фасета значения признаков могут просто перечисляться по некоторому порядку или образовывать сложную иерархическую структуру, если существует соподчиненность выделенных признаков.

Плюсы: большая емкость системы и высокая степень гибкости, поскольку при необходимости можно вводить дополнительные фасеты и изменять их место в формуле. При изменении характера задач или характеристик объектов классификации разрабатываются новые фасеты или дополняются новыми признаками уже существующие фасеты без коренной перестройки структуры всего классификатора.

Минусы - сложность структуры и низкая степень заполненности системы.

В современных классификационных схемах часто одновременно используются оба метода классификации. Это снижает влияние недостатков методов классификации и расширяет возможность использования классификаторов в информационном обеспечении управления.

Рассмотренные системы классификации хорошо приспособлены для организации поиска с целью последующей обработки информации, но лишь частично решают проблему содержательного поиска информации при принятии управленческих решений.

#### Дескрипторная система классификации

Для поиска показателей и документов по набору содержательных признаков используется информационный язык дескрипторного типа, который характеризуется совокупностью терминов (дескрипторов) и набором отношений между терминами.

Содержание документов или показателей можно достаточно полно и точно отразить с помощью списка ключевых слов — дескрипторов. **Дескриптор** — это термин естественного языка (слово или словосочетание), используемый при описании документов или показателей, который имеет самостоятельный смысл и неделим без изменения своего значения.

Для того чтобы обеспечить точность и однозначность поиска с помощью дескрипторного языка, необходимо предварительно определить все постоянные отношения между терминами:

* родовидовые (включение классов),
* отношения синонимии, омонимии и полисемии,
* ассоциативные отношения (объединение по общим свойствам).

Все выделенные отношения явно описываются в систематическом словаре понятий — **тезаурусе**, который разрабатывается с целью проведения индексирования документов, показателей и информационных запросов.

## Кодирование технико-экономической информации

Для полной формализации информации недостаточно простой классификации, поэтому проводят следующую процедуру — кодирование.

**Кодирование** — это процесс присвоения условных обозначений объектам и классификационным группам по соответствующей системе кодирования. Кодирование реализует перевод информации, выраженной одной системой знаков, в другую систему, то есть перевод записи на естественном языке в запись с помощью кодов.

**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 и смешанный.

Код характеризуется следующими параметрами:

* длиной;
* основанием кодирования;
* структурой кода, под которой понимают распределение знаков по признакам и объектам классификации;
* степенью информативности, рассчитываемой как частное от деления общего количества признаков на длину кода;
* коэффициентом избыточности, который определяется как отношение максимального количества объектов к фактическому количеству объектов.

К методам кодирования предъявляются определенные требования:

* код должен осуществлять идентификацию объекта в пределах заданного множества объектов классификации;
* желательно предусматривать использование в качестве алфавита кода десятичных цифр и букв;
* необходимо обеспечивать по возможности минимальную длину кода и достаточный резерв незанятых позиций для кодирования новых объектов без нарушения структуры классификатора.

Методы кодирования могут носить самостоятельный характер – регистрационные методы кодирования, или быть основанными на предварительной классификации объектов – классификационные методы кодирования.

#### Регистрационные методы кодирования

**Регистрационные методы** кодирования бывают двух видов:

1. порядковый

Кодами служат числа натурального ряда. Каждый из объектов классифицируемого множества кодируется путем присвоения ему текущего порядкового номера.

Плюсы: обеспечивает довольно большую долговечность классификатора при незначительной избыточности кода; обладает наибольшей простотой; использует наиболее короткие коды и лучше обеспечивает однозначность каждого объекта классификации; обеспечивает наиболее простое присвоение кодов новым объектам, появляющимся в процессе ведения классификатора.

Минус: отсутствие в коде какой-либо конкретной информации о свойствах объекта, а также сложность машинной обработки информации при получении итогов по группе объектов классификации с одинаковыми признаками.

1. серийно-порядковый

Кодами служат числа натурального ряда с закреплением отдельных серий этих чисел (интервалов натурального ряда) за объектами классификации с одинаковыми признаками. В каждой серии, кроме кодов имеющихся объектов классификации, предусматривается определенное количество кодов для резерва.

#### Классификационные коды

**Классификационные коды** используют для отражения классификационных взаимосвязей объектов и группировок и применяются в основном для сложной логической обработки экономической информации. Можно разделить на две подгруппы в зависимости от того, какую систему классификации используют для упорядочения объектов

1. **Последовательные системы** кодирования характеризуются тем, что они базируются на предварительной классификации по иерархической системе. Код объекта классификации образуется с использованием кодов последовательно расположенных подчиненных группировок, полученных при иерархическом методе кодирования. В этом случае код нижестоящей группировки образуется путем добавления соответствующего количества разрядов к коду вышестоящей группировки.
2. **Параллельные системы** кодирования характеризуются тем, что они строятся на основе использования фасетной системы классификации и коды группировок по фасетам формируются независимо друг от друга.

В параллельной системе кодирования возможны два варианта записи кодов объекта:

1. Каждый фасет и признак внутри фасета имеют свои коды, которые включаются в состав кода объекта. Такой способ записи удобно применять тогда, когда объекты характеризуются неодинаковым набором признаков. При формировании кода какого-либо объекта берутся только необходимые признаки.
2. Для определения групп объектов выделяется фиксированный набор признаков и устанавливается стабильный порядок их следования, то есть устанавливается фасетная формула. В этом случае не надо каждый раз указывать, значение какого из признаков приведено в определенных разрядах кода объекта.

Плюсы: Гибкость структуры кода, обусловленную независимостью признаков, из кодов которых строится код объекта классификации. При решении конкретных задач можно использовать коды только тех необходимых признаков объектов (сокращение кода). Группировку объектов по любому сочетанию признаков. Хорошо приспособлен для машинной обработки информации. По конкретной кодовой комбинации легко узнать, набором каких характеристик обладает рассматриваемый объект. При этом из небольшого числа признаков можно образовать большое число кодовых комбинаций. Набор признаков при необходимости может легко пополняться присоединением кода нового признака.

### Рекомендации по выбору методов классификации:

Основой классификатора должны быть наиболее существенные признаки классификации, соответствующие характеру решаемых с помощью классификатора задач.

При соподчиненных признаках классификации и стабильном комплексе задач, для решения которых предназначен классификатор, целесообразно использовать иерархический метод классификации. Критерием выбора последовательности использования признаков классификации по ступеням классификации является статистика запросов к классификатору. На верхних ступенях классификации в классификаторе должны использоваться признаки с наибольшим числом запросов. По этой же причине на верхних ступенях классификации выбирают наименьшее основание кода.

При несоподчиненных признаках классификации и при большой динамичности решаемых задач целесообразно использовать фасетный метод классификации.

## Общероссийский классификатор предприятий и организаций

Рассмотрим применение идентификационного метода кодирования на примере известный всем кода ОКПО (ОК 007–93). ОКПО - Общероссийский классификатор предприятий и организаций является, по сути, главным государственным классификатором хозяйствующих субъектов Российской Федерации.

ОКПО обеспечивает достоверность сведений относительно идентификации хозяйствующего субъекта РФ. Исходя из этого, отличающийся постоянством код является нерушимой базой для всех государственных информационных систем Росстата.

В качестве единого идентификатора ОКПО может быть использован при межведомственном обмене информацией, а также для свободной интеграции различных федеральных баз данных и иных ресурсов о хозяйствующих субъектах информационного характера.

### Структура и разделы классификатора ОКПО

Классификатор ОКПО включает следующие разделы:

І раздел содержит важную информацию о хозяйствующих субъектах, к которым причисляются юридические лица, а также их филиалы и представительства, организации, ведущие деятельность без получения статуса юридического лица.

ІІ раздел заключает в себе сведения общего порядка о хозяйствующих субъектах – индивидуальных предпринимателях (сокращенно – ИП).

С точки зрения структуры классификатора, оба раздела ОКПО состоят из:

* блока классификационных признаков;
* блока наименования объектов;
* блока идентификации.



Структура классификатора ОКПО

#### Блок идентификации

В блоке идентификации ОКПО значится сам код ОКПО, составленный путем кодирования.

Он насчитывает строго восемь цифровых символов, семь из которых содержат в себе идентификатор, а последний, восьмой знак – контрольное число.

Два первых числа кода определяются кодом отрасли, следующие пять – регистрационным номером предприятия. Т.о. использовано серийно-порядковое кодирование, серии формируются по отраслям.

#### Блок классификации

Блок классификации ОКПО также представляет код, построенный с использованием серийно-порядкового метода кодирования. Однако, в этом случае, кодировка состоит уже из десяти цифровых символов, девять из которых являются кодом классификации, а десятый, заключительный знак – контрольное число. Это код построен на основе иерархической системы классификации.

В этом блоке используются кодировки следующих общероссийских классификаторов:

* ОКОГУ – группировочный код общероссийского классификатора органов государственной власти и управления, который характеризует ведомственную принадлежность, а также организационную и административную подчиненность всех хозяйствующих субъектов или для тех из них, которые не имеют вышестоящего органа управления;
* ОКАТО – группировочный 11-разрядный код общероссийского классификатора объектов, подлежащих административно-территориальному делению, который указывает на территориальное местонахождение хозяйствующего субъекта;
* ОКТМО – кодировка общероссийского классификатора территории муниципального значения, на которой находится его хозяйствующий субъект;
* ОКВЭД – код общероссийского классификатора, указывающий на вид экономической деятельности, которая максимально точно может охарактеризовать хозяйствующий субъект;
* ОКФС – кодировка общероссийского классификатора, указывающая на формы собственности хозяйствующего субъекта;
* ОКОПФ – код общероссийского классификатора с учетом организационно-правовых форм хозяйствующего субъекта.

#### Блок наименования объектов

В этом блокесодержится информация о полном и сокращенном наименовании хозяйствующего субъекта на государственном русском языке, а при наличии названия субъекта на иностранном языке название вносится дополнительной строкой на английском языке.

## Методика расчета контрольного числа для всех кодов статистики

Контрольное число рассчитывается, исходя из Методики расчета контрольного числа, которая приводится в Правилах стандартизации ПР 50.1.024-2005 «Основные положения и порядок осуществления работ по разработке, ведению и применению общероссийских классификаторов».

Контрольное число рассчитывается следующим образом:

* Разрядам кода в общероссийском классификаторе, начиная со старшего разряда, присваивается набор весов, соответствующий натуральному ряду чисел от 1 до 10. Если разрядность кода больше 10, то набор весов повторяется.
* Каждая цифра кода умножается на вес разряда и вычисляется сумма полученных произведений.
* Контрольное число для кода представляет собой остаток от деления полученной суммы на модуль «11».
* Контрольное число должно иметь один разряд, значение которого находится в пределах от 0 до 9. Если получается остаток, равный 10, то для обеспечения одноразрядного контрольного числа необходимо провести повторный расчет, применяя вторую последовательность весов, сдвинутую на два разряда влево (3, 4, 5,…). Если в случае повторного расчета остаток от деления вновь сохраняется равным 10, то значение контрольного числа проставляется равным «0».

Пример расчета контрольного числа для шестизначного кода 563921:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Вес разряда | 1 | 2 | 3 | 4 | 5 | 6 |
| Значение разряда кода | 5 | 6 | 3 | 9 | 2 | 1 |

Сумма произведений, полученных путем умножения каждой цифры кода на вес разряда, равна 78. Остаток от деления полученной суммы на модуль «11» равен 1 (78=7\*11+1). Контрольное число для данного кода равно 1.