

Практика 3_часть 2

Классификаторы в ИТ:
разработка системного
классификатора

1_Состав ИО ИС

Внемашинное ИО

- *классификаторы* технико-экономической информации,
- унифицированная система документации,
- методические инструктивные материалы

Машинное ИО

- макеты/экранные формы для ввода первичных данных в ЭВМ или вывода результатной информации,
- структура *информационной базы*: входные, выходные файлы, база данных

2_Системы классификации информации

Классификация — это разделение множества объектов на подмножества по их сходству или различию в соответствии с принятыми методами.

Классификация фиксирует закономерные связи между классами объектов. Под **объектом** понимается любой предмет, процесс, явление материального или нематериального свойства.

Объекты классификации в ИС: справочные реквизиты-признаки; наименование показателей и документов; наименование компонентов проекта ИС.

Система классификации позволяет сгруппировать объекты и выделить определенные классы, которые будут характеризоваться рядом общих свойств.

Совокупность правил распределения объектов множества на подмножества называется **системой классификации**.

3_Системы классификации информации

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

Например, признак «роль предприятия-партнера в отношении деятельности объекта автоматизации» позволяет разделить все предприятия на две группы (на два подмножества): «поставщики» и «потребители».

Классификационная группировка — это множество или подмножество, объединяющее часть объектов классификации по одному или нескольким признакам.

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

4_Виды классификаторов

Международные классификаторы входят в состав Системы международных стандартов и обязательны для передачи информации между организациями разных стран мирового сообщества.

Общегосударственные (общесистемные) классификаторы, обязательны для организации процессов передачи и обработки информации между экономическими системами государственного уровня внутри страны.

Отраслевые классификаторы используют для выполнения процедур обработки информации и передачи ее между организациями внутри отрасли.

Локальные классификаторы используют в пределах отдельных предприятий.

Системные классификаторы, принятые отдельным предприятием для применения в рамках своей автоматизированной системы. Они содержат информацию, необходимую для решения задач в конкретной АС и отсутствующую в национальном или отраслевом классификаторе.

5_Свойства систем классификации информации

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

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

Степень заполненности системы определяется как частное от деления фактического количества группировок на величину емкости системы.

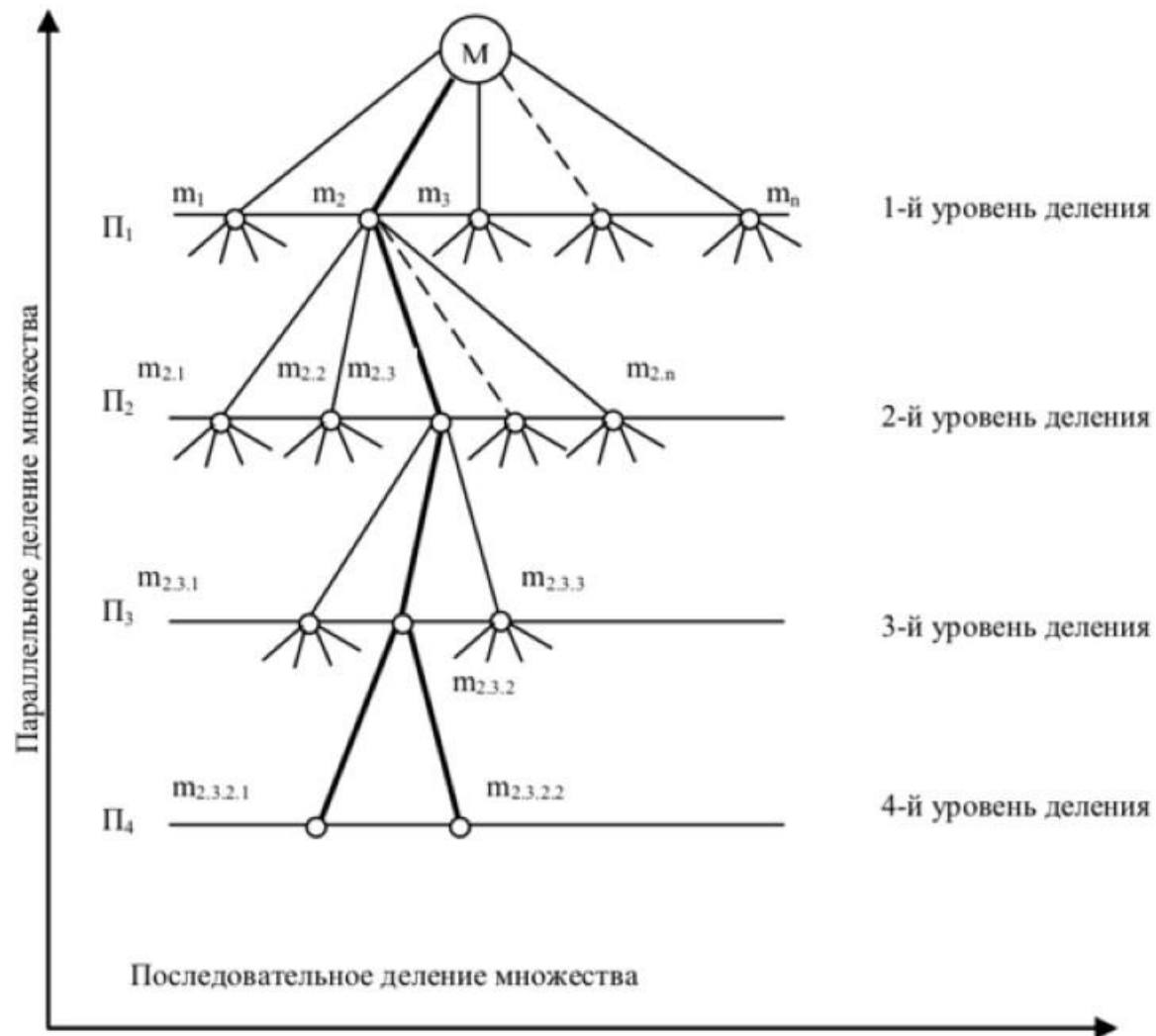
6_Иерархическая система классификации

Иерархический метод: последовательное разделение множества объектов на подчиненные, зависимые классификационные группировки.

При построении иерархической системы классификации сначала выделяется некоторое *множество объектов*, подлежащее классифицированию, для которого определяются *полное множество признаков классификации и их соподчиненность друг другу*, затем производится *разбиение исходного множества объектов на классификационные группировки на каждой ступени 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;
- количество **уровней** (ступеней) классификации определяет **глубину** классификации.

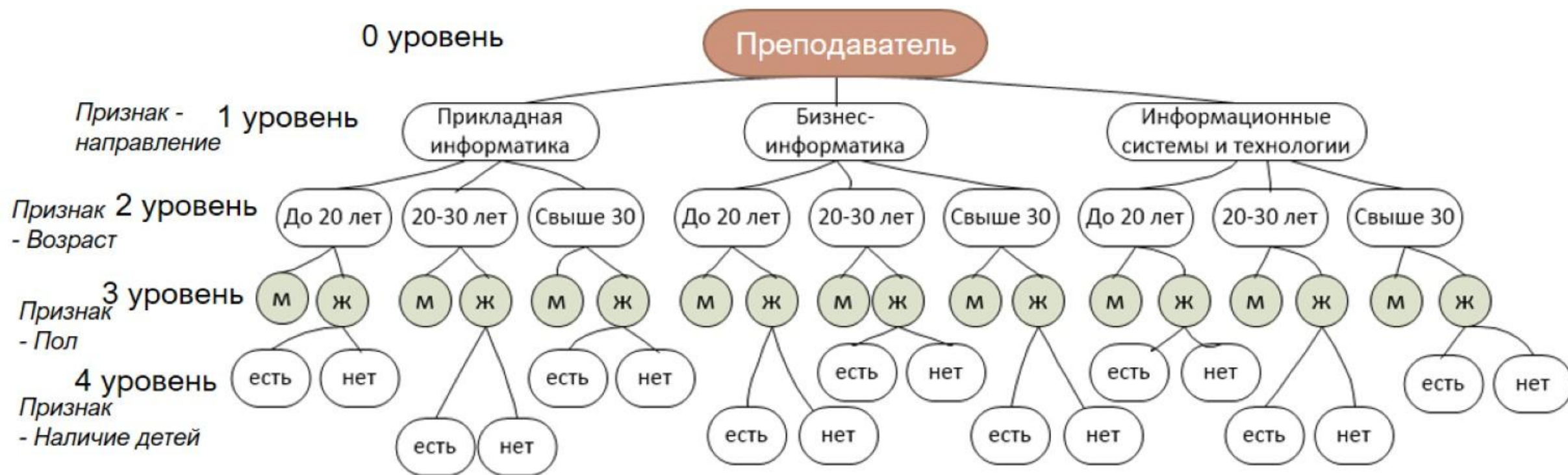


7_Схема иерархической классификации

8_ Построение иерархического классификатора

1. Выделение множества объектов, подлежащего классификации.
2. Определение полного множества признаков классификации и их соподчиненности друг другу.
3. Разбиение исходного множества объектов на классификационные группировки на каждом уровне классификации в соответствии с признаком:
 - исходное множество элементов составляет 0-й уровень и делится в зависимости от выбранного классификационного признака на классы (группировки), которые образуют 1-й уровень;
 - каждый класс 1-го уровня в соответствии со своим, характерным для него классификационным признаком делится на подклассы, которые образуют 2-й уровень;
 - каждый класс 2-го уровня аналогично делится на группы, которые образуют 3-й уровень и т.д.

9_Пример иерархической классификации



10_ Требования к иерархической классификации

- 1. Непересекаемость** классификационных группировок, расположенных на одной ступени классификации. Это означает, что классификационные группировки, расположенные на одной и той же ступени классификации, не должны включать аналогичных понятий.
- 2.** Для разделения любой классификационной группировки на подчиненные группировки должен использоваться только один признак.
- 3. Логичность и последовательность деления группировок** на нижестоящие и полнота этого деления. В соответствии с этим требованием на *верхних ступенях классификации должны использоваться признаки, к которым в дальнейшем будет обращено наибольшее число запросов.* Полнота деления означает, что сумма подмножества всегда должна давать делимое множество объектов, не должна оставаться какая-то часть объектов, не вошедшая в состав классификационных группировок

11_Преимущества и недостатки иерархической классификации



+ Логичность



+ Простота построения



+ Неограниченная емкость



- Жесткость классификационной схемы

12_Фасетная система классификации

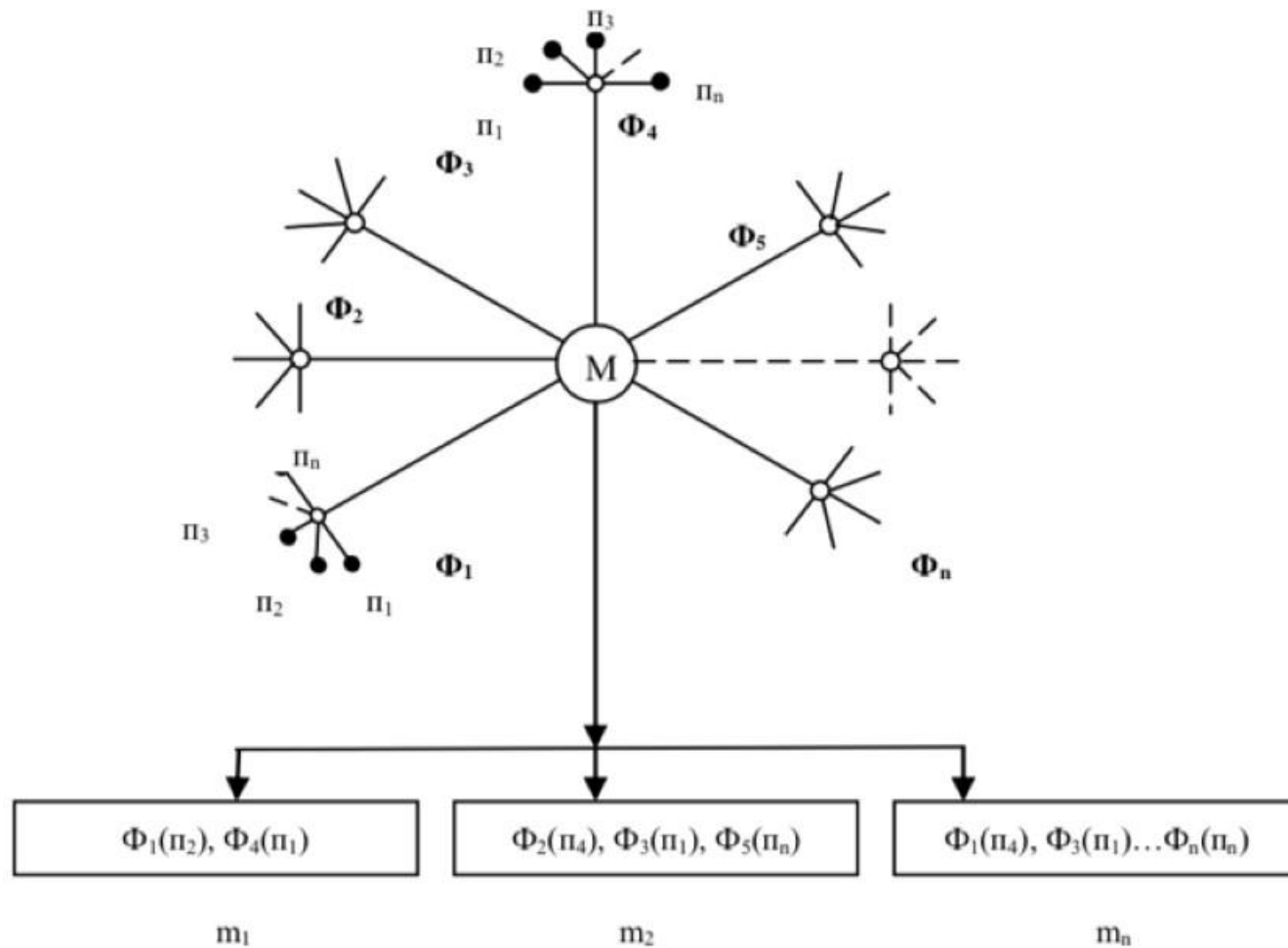
Под фасетным методом классификации понимается **параллельное** разделение множества объектов на независимые классификационные группировки. Фасетный метод классификации в отличие от иерархического позволяет выбирать признаки классификации **независимо** – как друг от друга, так и от **семантического** содержания классифицируемого объекта.

Разрабатывается система таблиц признаков объектов классификации, называемых фасетами.

Фасет (facet — рамка) – признак классификации. Каждый фасет содержит совокупность однородных значений данного классификационного признака. Причем значения в фасете могут располагаться в произвольном порядке, хотя *предпочтительнее их упорядочение*.

13_Схема фасетной классификации

Для каждого **объекта** задается конкретная группировка **фасетов** структурной формулой, в которой отражается их порядок следования



Фасеты						
Φ_1	Φ_2	Φ_3	...	Φ_n	...	Φ_i

14_ Построение фасетного классификатора и требования

1. Выделение некоторого множества объектов, подлежащего классификации в соответствии с решаемой задачей.
2. Выделение значимых признаков объекта.
3. Разработка системы таблиц признаков объектов классификации (фасетов).

Требования к фасетному классификатору:

1. Должен соблюдаться **принцип непересекаемости фасета**, то есть состав признаков одного фасета не должен повторяться в других фасетах этого же класса;
2. В состав классификатора должны быть включены только такие фасеты и признаки, которые *необходимы для решения конкретных задач*.

Направления	Возраст	Пол	Дети
ПИ	до 20 лет	м	есть
БИ	20-30 лет	ж	нет
ИСиТ	свыше 30 лет		
...			

15_Пример фасетной классификации

16_Преимущества и недостатки фасетной классификации



+ Гибкость классификационной схемы



- Низкая заполненность системы



+ Неограниченная емкость



- Сложность структуры

17_Системы кодирования

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 и смешанный.

18_Требования к коду

Код характеризуется следующими параметрами:

- длиной — число позиций в коде;
- структурой — порядок расположения в коде символов, используемых для обозначения классификационного признака;
- степенью информативности – отношение общего количества признаков к длине кода;
- коэффициентом избыточности – отношение фактического количества объектов к максимально возможному 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*.

19_Методы кодирования

1. Классификационное кодирование

проведение предварительной классификации объектов:

- **последовательное** кодирование (на основе иерархической классификации),
- **параллельное** кодирование (на основе фасетной классификации)

2. Регистрационное кодирование

без предварительной классификации объектов:

- **порядковое** кодирование
- **серийно-порядковое** кодирование

20_Классификационное кодирование: последовательное

Последовательный метод кодирования – метод, при котором код классификационной группировки и (или) объекта классификации образуется с использованием кодов *последовательно расположенных подчиненных группировок, полученных при иерархическом методе кодирования.*

Порядок последовательного кодирования:

1. Определить число классификационных признаков и их соподчиненность.
2. Установить число позиций в каждом классификационном признаке.
3. Произвести кодирование порядковыми номерами сначала старшего признака, затем следующих признаков внутри старших в соответствии с их разрядностью.
4. Составить классификатор.

21_Пример последовательного кодирования

Структура кода объекта «Преподаватель»:

XXXX, где X - значение десятичного разряда.

Содержание структуры кода, начиная со старшего разряда:

1-й разряд выделен для классификационного признака «направление» и имеет следующие значения: 1–Прикладная информатика; 2–Бизнес-информатика; 3–Информационные системы и технологии.

2-й разряд выделен для классификационного признака «возраст» и имеет следующие значения: 1–до 20 лет; 2–от 20 до 30 лет; 3–свыше 30 лет.

3-й разряд выделен для классификационного признака «пол» и имеет следующие значения: 1–мужчины; 2–женщины.

4-й разряд выделен для классификационного признака «наличие детей» и имеет следующие значения; 1–есть дети; 2–нет детей.

Пример кода: 1310 — преподаватели, работающие по направлению Прикладная информатика, свыше 30 лет, мужчины

22_Классификационное кодирование: параллельное

Параллельный метод кодирования – метод, при котором код классификационной группировки и (или) объекта классификации образуется с использованием независимых группировок, полученных при фасетном методе классификации.

При этом методе кодирования признаки объекта кодируются **независимо** друг от друга.

Для параллельного метода 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* и различны их числом. 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 признака приведено в определенных разрядах кода объекта.

23_Пример параллельного кодирования

Структура кода объекта «Преподаватель» при фиксированном наборе признаков (фасетная формула):

XXXX, где X – значение десятичного разряда.

Содержание структуры кода:

1-й (старший) разряд выделен для фасета «пол» и имеет следующие значения: 1–мужчины; 2–женщины.

2-й разряд выделен для фасета «наличие детей» и имеет следующие значения: 1–есть дети; 2–нет детей.

3-й разряд выделен для фасета «возраст» и имеет следующие значения: 1–до 20 лет; 2–от 20 до 30 лет; 3–свыше 30 лет.

4-й разряд выделен для фасета «наименование направления» и имеет следующие значения 1–Прикладная информатика, 2–Бизнес-информатика, 3–Информационные системы и технологии.

Пример кода: 1021 — мужчины возраста от 20 до 30 лет, являющиеся преподавателями направления ПИ.

24_Регистрационное кодирование

Порядковая система кодирования предполагает **последовательную** нумерацию объектов числами *натурального ряда*. В этом случае каждый из объектов классифицируемого множества кодируется путем присвоения ему текущего порядкового номера.

Серийно-порядковый метод кодирования – это такой *метод*, при котором 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*. Резерв кодов располагается в середине или в конце серии.

Серийно-порядковый метод кодирования целесообразно применять для объектов, **имеющих два соподчиненных признака**.

25_ Общие принципы классификации и кодирования ТЭИ

1. **Основой** классификатора должны быть наиболее **существенные признаки классификации**, соответствующие **характеру** решаемых с помощью классификатора задач. При этом данные признаки могут быть или *соподчиненными* (т.е. каждый последующий зависит от предыдущего), или *несоподчиненными* (характеризуют объект с независимых друг от друга сторон).
2. При **соподчиненных** признаках классификации и **стабильном комплексе задач**, для решения которых предназначен классификатор, целесообразно использовать *иерархический* метод классификации, который представляет собой последовательное разделение множества объектов на подчиненные классификационные группировки.
3. При **несоподчиненных** признаках классификации и при **большой динамичности решаемых задач** целесообразно использовать *фасетный* метод классификации.
4. Важным вопросом является правильный выбор *последовательности использования признаков классификации* по ступеням классификации при *иерархическом* методе классификации. Критерием при этом является **статистика запросов к классификатору**. В соответствии с этим критерием на верхних ступенях классификации в классификаторе должны использоваться признаки, к которым будут наиболее частые запросы.