

И-Л5-04.10.2025(18.10.2025)

#лекция

Кодирование при классификации информации

▣ Определение

Классификатор (классификационная схема) -

систематизированный перечень наименования объектов, каждым из которых в соответствии дан уникальный код.

Классификация объектов производится согласно правилам распределения заданного множества объектов на подмножества (классификационные группировки) в соответствие установленными признаки их различия или сходства.

Применяются в автоматизированных системах управления и области обработки информации.

Классификатор является стандартным кодовым языком унифицированных документов при обработке в автоматизированных системах. Классификаторы разрабатываются как на уровне отдельных предприятий (организаций), так и на уровне государств.

Существуют следующие уровни классификаторов:

- *международные* классификаторы (стандартные классификаторы, используемые по всему миру)
- *межгосударственные* (классификаторы, используемые в рамках экономических союзов и других межгосударственных объединений)
- *межотраслевые*
- *отраслевые*

- *региональные*
 - *системные* (используются отдельной организацией для применения в рамках своей организации. Они содержат информации, необходимую для решения задач в конкретной автоматизированной системе и отсутствующую в вышестоящих классификаторах).
-

Методы классификаций

1. *Иерархический*
2. *Фасетный (классификация двоеточия)*
3. *Дескрипторный (описательный)*

Под иерархическим методом классификации понимается метод, при котором заданное множество последовательно делится на подчинённые подмножества, постепенно конкретизируя объект классификации, при этом основанием деления служит *некоторый выбранный признак*.

Совокупность получившихся группировок при этом образует иерархическую древовидную структуру в виде ветвящегося графа, узлами которого являются группировки.

Требования к иерархическому классификатору:

1. классификационные группировки, расположенные на одной ступени классификатора *не должны пересекаться*, то есть не должны включать в себя аналогичных понятий;
2. на каждой ступени классификатора для разделения вышестоящей группировки должен использоваться только *один признак*;
3. сумма подмножества всегда должна давать *делимое множество* объектов;
4. не должно оставаться объектов *не вошедших в состав* классификационной группировки.

Основными преимуществами иерархического метода является:

- *большая информационная ёмкость;*
- *традиционность и привычность применения;*
- *возможность создания для объектов классификации мнемонических кодов, несущих смысловую нагрузку.*

Значительным **недостатком** иерархической классификации является *слабая гибкость структуры*, обусловленная фиксированным основанием деления и заранее установленным порядком следования, не допускающим включение новых объектов и классификационных группировок.

Таким образом, при изменении состава объектов классификации и характеристик с помощью классификационных задач требуется *коренная переработка* всей классификационной схемы.

Фасетная классификация - множество, элементами которого является множество.

Совокупность нескольких независимых классификаций, осуществляемых одновременно по различным основаниям, в которой:

1. понятия представлены в виде пересечения ряда признаков;
2. классификационные индексы синтезируются посредством комбинирования фасетных признаков в соответствии с фасетной формулой;

Основой классификации является привычной нам отнесение объекта к разным категориям. Фасетный метод классификации подразумевает параллельное разделение множества объектов на независимые классификационные группировки. При этом не предполагается жесткой классификационной структуры и заранее построенных конечных группировок. Классификационные группировки образуются путём комбинации значений, взятых из соответствующих фасетов.

Последовательность расположения фасетов при образовании классификационной группировки задаётся фасетной формулой. Количество фасетных формул определяется возможными сочетаниями признаков.

К классификатору, построенному на фасетном методе предъявляются следующие требования:

1. должен соблюдаться *принцип непересекаемости фасета*, то есть состав признаков одного фасета не должен повторяться в других фасетах этого же класса
 2. В состав классификаторов должны быть включены только такие фасеты и признаки, необходимые для решения задач.
-

Для организации поиска информации, для введения тезауруса (словарей) используется дескрипторная система классификации, язык которой приближается к естественному языку описания информационных объектов. Особенно широко она используется в библиотечной системе поиска.

Суть дескрипторного метода:

- отбирается совокупность ключевых слов или словосочетаний, описывающих определённую предметную область или совокупность однородных объектов.
- Выбранные ключевые слова и словосочетания подвергаются нормализации, то есть из совокупности синонимов выбирается один или несколько наиболее употребимых.
- создается словарь дескрипторов, то есть словарь ключевых слов и словосочетаний, отобранных в результате процедуры нормализации.
- между дескрипторами устанавливаются связи, которые позволяют расширить область поиска информации.

- синонимические связи, указывающие на некоторую совокупность ключевых слов, как синонимов.
- родовидовые связи, отображающий включение некоторого класса объектов в более высокий класс
- ассоциативные, соединяющие дескрипторы, обладающие общими свойствами

Методы кодирования в классификаторах

Под кодированием понимается присвоение кода классификационной группировки или объекту классификации. Кодирование предназначено для формализованного описания наименований, различных аспектов данных.

Обычно кодирование представляет собой процесс обозначения исходного множества объектов или сообщений, набором символов заданного алфавита на основе совокупности определённых правил:

- порядковый

Каждый из объектов множества кодируется с помощью; обеспечивает простоту добавления новых объектов и краткость кода, однако такой код не несёт никакой информации об объекте

- серийно-порядковый

Кодами служат числа натурального ряда с закреплённой отдельной серией этих чисел за объектами классификации с одинаковыми признаками.

- последовательный метод

В кодовом обозначении знаки на каждой ступени деления зависят от результатов разбиения на предыдущих ступенях. В результате кодовое обозначение группировки даёт информацию о последовательности признаков, характеризующих эту группировку. Наиболее подходит иерархическому методу.

- параллельный метод

Признаки классификации кодируются независимо друг от друга

определенными разрядами или группой разрядов кодового обозначения. Метод параллельного кодирования чаще всего используется при фасетной классификации, но также применяется в иерархической классификации.