ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 55021— 2012/ ISO/IEC TR 9789:1994

Информационная технология

РУКОВОДСТВО ПО ОРГАНИЗАЦИИ И ПРЕДСТАВЛЕНИЮ ЭЛЕМЕНТОВ ДАННЫХ ПРИ ОБМЕНЕ ДАННЫМИ

Методы и принципы кодирования

(ISO/IEC TR 9789:1994, IDT)

Издание официальное

Москва
Стандартинформ
2018

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Федеральным государственным унитарным предприятием «Государственный научно-исследовательский и конструкторско-технологический институт «ТЕСТ» (ФГУП ГосНИИ «ТЕСТ») на основе собственного перевода на русский язык англоязычной версии международного документа, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 22 «Информационные технологии»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 24 сентября 2012 г. № 417-ст

4 Настоящий стандарт идентичен международному документу ISO/IEC TR 9789:1994 «Информационная технология. Руководство по организации и представлению элементов данных при обмене данными. Методы и принципы кодирования» (ISO/IEC TR 9798:1994 «Information technology — Guidelines for the organization and representation of data elements for data interchange — Coding methods and principles», IDT).

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им национальные стандарты, сведения о которых приведены в дополнительном приложении ДА

5 ВВЕДЕН ВПЕРВЫЕ

6 ПЕРЕИЗДАНИЕ. Декабрь 2018 г.

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© ISO, 1994 — Все права сохраняются
© Стандартинформ, оформление, 2013, 2018

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Принципы кодирования 3
4.1 Информация и кодирование 3
4.2 Кодирование 3
5 Цели кодирования 4
5.1 Идентификация 4
5.2 Классификация 4
5.3 Ключ к дальнейшей информации 5
6 Типы кодов 5
6.1 Формы кодов 5
6.2 Последовательные коды 6
6.3 Случайные коды 8
6.4 Мнемонические коды 8
6.5 Коды, основанные на сокращениях 9
6.6 Матричные коды 10
6.7 Иерархические коды 10
6.8 Сопоставляющие коды 11
6.9 Комбинированные коды 12
6.10 Коды, суммирующие значения 12
7 Характеристики кодов 13
7.1 Общие характеристики 13
7.2 Однозначность 13
7.3 Расширяемость 13
7.4 Компактность 14
7.5 Простота 14
7.6 Многосторонность 14
7.7 Пригодность для сортировки 14
7.8 Стабильность 14
7.9 Выразительность 14
7.10 Размер 14
7.11 Структура и формат 15
7.12 Мощность 15
8 Представление закодированных значений 16
8.1 Наборы символов 16
8.2 Способы представления 17
9 Проектирование кодов 17
9.1 Общие положения 17
9.2 Количество символов 18
9.3 Формат закодированных представлений 18
9.4 Наборы символов 19
9.5 Соглашения о присваивании 19

9.6 Выявление ошибок с помощью контрольных символов 20
9.7 Рекомендации для некоторых прикладных областей 20
9.8 Практические рекомендации 21
Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов национальным стандартам 23

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Информационная технология

РУКОВОДСТВО ПО ОРГАНИЗАЦИИ И ПРЕДСТАВЛЕНИЮ ЭЛЕМЕНТОВ ДАННЫХ ПРИ ОБМЕНЕ ДАННЫМИ

Методы и принципы кодирования

Information technology. Guidelines for the organization and representation of data elements for data interchange. Coding methods and principles

Дата введения — 2014—01—01

1 Область применения

Настоящий стандарт представляет общее руководство по способам, которыми данные могут быть выражены в виде кодов.

В стандарте описаны цели кодирования, характеристики, преимущества и недостатки различных методов кодирования, особенности кодов и дано руководство по проектированию кодов.

Настоящий стандарт не предназначен непосредственно для какой-либо конкретной прикладной области и не зависит от какого-либо конкретного метода проектирования прикладных систем или обмена данными.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ISO/IEC 646:1991, Information technology — ISO 7-bit coded character set for information interchange (Информационные технологии. 7-битный набор кодированных символов ISO для обмена информацией)

ISO 2375:1985[1)], Data processing — Procedure for registration of escape sequences (Обработка данных. Процедура регистрации выходной последовательности)

ISO 2382-4:1987[2)], Information processing systems — Vocabulary — Part 4: Organization of data Bilingual edition (Системы обработки информации. Словарь. Часть 4. Организация данных)

ISO 3166:1993[3)], Codes for the representation of names of countries (Коды для представления названий стран)

ISO 7064:1983[4)], Data processing; Check character systems (Обработка информации. Системы контрольных символов)

ISO 8601:1988[5)], Data elements and interchange formats — Information interchange — Representation of dates and times (Элементы данных и форматы для обмена информацией. Обмен информацией. Представление дат и времени)

---

1) Заменен на ИСО/МЭК 2375:2003.

2) Заменен на ИСО/МЭК 2382-4:2015.

3) Заменен на ИСО 3166-1:2013.

4) Заменен на ИСО/МЭК 7064:2003.

5) Заменен на ИСО 8601:2004.

Издание официальное

ISO 9735:1988, Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules (Электронный обмен данными в управлении, торговле и на транспорте (EDIFACT). Синтаксические правила для прикладного уровня)

ISO/IEC 11179-3:1994[1)], Information technology. Specification and standardization of data elements. Part 3: Basic attributes of data elements (Информационные технологии. Спецификация и стандартизация элементов данных. Часть 3. Базовые атрибуты элементов данных)

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1 **атрибут** (attribute): Характеристика объекта.

3.2 **набор символов** (character set): Конечное множество различных символов, которое является полным для заданной цели.

Пример — Международная базовая версия набора символов ИСО 646.

3.3 **код** (code): Совокупность правил, которые отображают элементы одного набора в элементы другого набора.

Примечания

1 Элементы могут быть символами или строками символов.

2 Первый набор называется кодируемым, а второй — закодированным набором элементов.

3 Элемент закодированного набора может относиться к нескольким кодируемым.

3.4 **закодированный элемент** (code element): Результат применения *кода* к элементу в *кодируемом наборе*.

Примеры

1 «CDG» является представлением парижского аэропорта имени Шарля де Голля (Paris Charles de Gaulle) в коде для трехбуквенного представления названий аэропортов.

2 Семь двоичных цифр, представляющих удаление символа в ИСО 646.

3.5 **закодированный набор элементов** (code element set): Результат применения *кода* ко всем элементам *кодируемого набора*.

Пример — Все трехбуквенные международные представления названий аэропортов.

3.6 **закодированный набор** (code set): Синоним **закодированного набора элементов**.

3.7 **закодированное значение** (code value): Синоним **закодированного элемента**.

3.8 **закодированное представление** (coded representation): Синоним **закодированного элемента**.

3.9 **кодируемый набор** (coded set): Набор элементов, который отображается в другой набор в соответствии с *кодом*.

Пример — Перечень названий аэропортов, который отображается в соответствующий набор трехбуквенных сокращений.

3.10 **схема кодирования** (coding scheme): Синоним **кода**.

3.11 **данные кода** (data code): См. предпочтительный термин **закодированный элемент**.

3.12 **экземпляр элемента данных** (data element instance): Экземпляр *типа элемента данных*.

3.13 **тип элемента данных** (data element type): Категория данных, которая представляет понятие и свойства которой выражаются как набор атрибутов типа элемента данных, допускающий обеспечение информационного обмена.

3.14 **разделитель** (delimiter): Один или несколько символов, используемых для указания начала или окончания строки символов.

3.15 **сущность** (entity): Любое конкретное или абстрактное рассматриваемое понятие, включая ассоциации между самими понятиями.

3.16 **поле** (field): Специфицированная область в среде данных или в памяти, используемая для конкретного класса элементов данных.

[1)] Заменен на ИСО 11179-3:2003.

3.17 **идентификатор** (identifier): Один или несколько символов, используемых для идентификации или наименования типа элемента данных и, возможно, для указания некоторых его свойств.

3.18 **ключ** (key): Идентификатор в наборе типов элементов данных.

3.19 **позиция** (position): Любое положение в *строке*, которое может быть занято элементом и которое идентифицировано порядковым номером.

3.20 **строка** (string): Рассматриваемая как единое целое последовательность элементов одной и той же природы, таких как символы.

3.21 **таблица** (table): Упорядоченные данные, каждый элемент которых может быть идентифицирован с помощью аргументов или ключей.

4 Принципы кодирования

4.1 Информация и кодирование

В повседневной жизни информация понимается как факты и утверждения о рассматриваемых конкретных или абстрактных объектах, выраженные: данными, сообщениями, последующими уточнениями. Информация необходима для надлежащего выполнения любой возможной задачи, будь то в управлении, торговле, транспорте, науке и т. д. Точная, объективная и недвусмысленная информация является предварительным требованием в компьютерных информационных системах и при обмене данными между ними.

В формальных информационных системах данные выражены символами (обычно цифрами и буквами), арифметическими числами и описаниями, которые должны иметь фиксированный, неизменный смысл для каждого участника и, тем самым должны быть пригодными для целей обработки и коммуникации.

Пользователи, независимо от их функций и задач, должны быть способны корректно понимать, интерпретировать и обрабатывать свою информацию. Информация, совместно используемая разными группами пользователей или прикладными системами, должна иметь согласованное определение, т.е. семантическое значение понятия (коннотацию), все экземпляры понятия (денотацию) и согласованное представление.

Кодирование охватывает способы и формы выражения данных. Необходимо установить четкие соглашения об этих представлениях. Объяснение форм представления и закодированных элементов является частью спецификации данных.

4.2 Кодирование

Под кодированием понимается основанное на правилах упорядоченное присваивание закодированных элементов поименованному и определенному набору элементов. Кодирование осуществляется главным образом с помощью символов (обычно цифр и букв) и приводит к сжатому представлению.

Пример — Присваивание закодированного элемента «CDG» для краткого представления названия аэропорта «Paris Charles de Gaulle» (Парижский аэропорт имени Шарля де Голля). Это название аэропорта относится к перечню названий аэропортов, поддерживаемому Международной ассоциацией воздушного транспорта (International Air Transport Association — IATA). Эта организация имеет свод правил для установления кратких представлений.

Кодирование является необходимым инструментом для обработки информации. Кодирование информации позволяет людям и машинам ее записывать, интерпретировать, обрабатывать и передавать. Все виды информации могут быть закодированы: информация о продуктах, людях, процессах, документах, странах, валютах, грузах и т. п.

До выработки соглашений о методологии кодирования, т. е. о представлении информации, относящейся к событиям, действиям, конкретным или абстрактным объектам реального мира, необходимо исследовать, какие данные относятся к намечаемым приложениям. Анализ информации рассматриваемой области применения должен определить роль данных в структуре информации. В ходе анализа следует четко различать потребности идентификации, классификации и указания.

5 Цели кодирования

Информация о любом рассматриваемом абстрактном или конкретном объекте, действии или событии (их характеристики или атрибуты) может быть закодирована. Прежде чем принять соглашения о конфигурации их представления (правила кодирования), необходимо определить цели кодирования. Недостаточно просто спроектировать упорядоченное краткое представление некоторых данных. Сначала должны быть выяснены информационные требования.

В общем случае встречаются следующие требования:

- идентификация,
- классификация,
- ключ к последующей информации.

Когда применяется моделирование данных для спецификации прикладных систем или сообщений, используемых при обмене данными, цели пользователей в прикладной среде будут определять выбор тех категорий и атрибутов, которые должны быть приняты во внимание, а так же их взаимосвязи. От этих целей будут зависеть методы, которые должны использоваться для идентификации, классификации и указания.

5.1 Идентификация

Целью идентификации является отличие друг от друга элементов набора. Для того, чтобы это было возможно, прежде всего, следует определить, какие характеристики должны приниматься во внимание. На основе выбранных характеристик может быть проведено сравнение и оценено, равен ли данный элемент набора другому или отличается от него. С какой степенью подробности должны записываться характеристики для того, чтобы они указывали на подобие или различие элементов набора, зависит от прикладной области, для которой необходима идентификация.

Пример — Для контроля запасов канцелярских принадлежностей организация хочет идентифицировать различные типы листов бумаги. Иногда может оказаться достаточным различать форматы: А3, А4 или А5. В зависимости от использования бумаги могут быть добавлены другие характеристики, такие, как толщина, масса, химический состав. Если специфические требования не установлены, то может оказаться достаточным записывать формат. В других случаях, когда обработка и использование бумаги являются важными, должны быть указаны необходимые характеристики.

Таким образом, идентификация может быть определена как:

- систематическая регистрация характеристик элементов набора таким способом, что они могут отличаться друг от друга,
- характеристики, по которым должно проводиться различие, являются неотъемлемой частью рассматриваемого объекта или понятия. Увеличение степени детализации, которая должна отслеживаться, зависит от задач пользователя и области применения. Эти критерии приводят к планированию или выбору системы идентификации.

Примеры необходимой степени подробности

В реальном мире человек имеет имя и фамилию, родился в конкретной стране, в конкретный день, проживает в конкретной стране, имеет глаза конкретного цвета и т.д.

Приложение 1 — В государственном приложении желательно различать цвет глаз граждан.

Приложение 2 — В приложении для медицинских исследований желательно различать цвет глаз людей.

В этих приложениях требования степени различия не обязательно идентичны. В медицинском исследовательском приложении может требоваться различать большее количество цветов глаз, чем в государственном приложении, регистрирующем цвет глаз граждан.

На практике это может привести к компромиссу при выборе системы идентификации. Часто выбор системы определяется на основе желания иметь минимум систем идентификации для соответствия максимуму функций.

Цели и приложения определяют, какие внутренние характеристики должны рассматриваться.

5.2 Классификация

Целью классификации является группировка рассматриваемых объектов или понятий в классы в соответствии с предварительно определенными характеристиками на основе того, какое подобие мо-

жет быть установлено. Классификация часто используется для поддержки принятия решений или для выявления тенденций или направлений развития без рассмотрения каждого экземпляра множества по отдельности.

Таким образом, классификация может быть определена как:

- систематическое упорядочение элементов в группы или категории на основании сходства предварительно определенных характеристик;
- классификация осуществляется с помощью контрольных характеристик, т.е. тех характеристик, которые были приняты для рассматриваемого объекта или понятия или относятся к нему. Эти характеристики могут быть как внутренними, так и внешними.

Примеры контрольных характеристик

Место производства продукции, время оборачиваемости средств, сектор рынка, производственный процесс.

Информационные требования и бизнес-политика являются определяющими для выбора управляющих характеристик.

В организации могут применяться разные схемы классификации одного и того же типа объектов в зависимости от разных потребностей.

Пример

Продукт может быть классифицирован в соответствии с:

- ***функцией с точки зрения продажи,***
- ***процессом его производства,***
- ***ценностью с точки зрения инвентаризации,***
- ***массой/объемом с точки зрения транспортировки,***
- ***типом с точки зрения потребителей или статистики.***

5.3 Ключ к дальнейшей информации

Ключ является идентификатором в наборе типов элементов данных. Ключ должен быть уникален в контексте приложения или обмена данными. Во многих прикладных системах в качестве ключа к дальнейшей информации необходим ссылочный номер. Сам ключ может быть бессмысленным, но он предоставляет доступ к нужным данным.

Примеры

1 Порядковый номер может открывать доступ стороне, которой был передан порядок данных, товаров или услуг.

2 Номер артикула может быть связан с описанием, ценой, производственным процессом, местом производства.

3 Табельный номер может быть связан с сотрудником, его именем и адресом, днем рождения, должностью, зарплатой.

Ссылочные номера могут быть идентифицирующими в одной прикладной области и классифицирующими — в другой.

6 Типы кодов

В данном разделе приведено описание основных методов кодирования. Это должно помочь при выборе подходящей структуры кода на основании конкретных прикладных требований и характера элементов в наборе, который должен быть закодирован. Здесь же приведены принципы и критерии, которые должны быть рассмотрены при оценке альтернативных структур кода, а так же достоинства и недостатки каждого метода кодирования.

Возможен выбор различных структур кода. Приведенная ниже информация должна помочь при выборе наилучшего метода.

6.1 Формы кодов

Рассмотренные в этом разделе методы кодирования перечислены ниже. Приведенный набор методов не является исчерпывающим, но включает в себя все значительные типы. Многие применяемые на практике структуры кодов часто являются комбинациями этих основных типов.

Незначащие коды

Последовательный:

- возрастающий последовательный;
- групповой последовательный;
- упорядоченный последовательный (хронологически, алфавитно).

Случайный

Значащие коды

Мнемонический

Основанный на сокращениях

Матричный

Иерархический

Сопоставляющий

Комбинированный

Суммирующий значения

6.2 Последовательные коды

6.2.1 Принципы

Элементам набора, которые должны быть закодированы, присваиваются номера, последовательно взятые из упорядоченного множества чисел. Эти номера обычно являются натуральными целыми числами (например, начиная с 1), но могут использоваться и алфавитные символы, например, AAA, AAB, AAC, ...

Последовательное присвоение закодированных элементов может быть основано на списках, построенных различными способами, например:

1 Список натуральных целых чисел, ограниченный количеством желательных доступных возможностей.

2 Списки чисел, упорядоченных на основании некоторого алгоритма, например, только четные числа или числа, кратные 10.

6.2.2 Использование

В основном последовательные коды используются как самодостаточные коды для идентификации или указания позиции или как часть составного кода, часто — в дополнение к классифицирующему коду.

Примечание — В числовых полях с фиксированным количеством позиций начальные нули должны использоваться тогда, когда это применимо, для заполнения поля до требуемого количества позиций.

Пример — В поле, определенном как трехзначное, «один» кодируется как 001, а «пятнадцать» — как 015.

6.2.3 Типы последовательных кодов

Имеется четыре типа последовательных кодов:

- возрастающий;
- абсолютно возрастающий;
- групповой;
- упорядоченный.

6.2.3.1 Возрастающий последовательный код

Элементам кодируемого набора присваиваются значения кода, определяемые увеличением предыдущего присвоенного закодированного элемента на заранее определенное число, например, 1 (абсолютно возрастающий), 2 (в случае четных номеров) или 10 (если присваиваются только кратные 10 номера).

При этом методе закодированное значение не выражает никакого смысла. Сходные элементы набора не группируются.

Основанием для присвоения закодированных значений, отличных от возрастающих на 1, может быть требование использовать промежуточные закодированные значения для последующих изменений исходного кодируемого набора.

Преимущества:

- быстрое присвоение закодированных значений;
- компактность;

- простота подтверждения допустимости закодированных представлений.

Недостатки:

- невозможна классификация или группировка элементов набора по закодированным представлениям;
- максимальная вместимость используется не полностью.

6.2.3.2 Абсолютно возрастающие последовательные коды

Элементам кодируемого набора присваиваются значения кода, определяемые увеличением предыдущего присвоенного закодированного элемента на 1.

При этом методе закодированное значение не выражает никакого смысла. Сходные элементы набора не группируются.

В дополнение к упомянутым в 6.2.3.1 преимуществам, используется максимальная вместимость кода.

6.2.3.3 Групповые последовательные коды

В этом типе кода категориям элементов набора, имеющим что-либо общее, присваиваются диапазоны закодированных значений.

До присвоения закодированного значения должна быть определена категория, к которой относится элемент набора. Затем присваивается следующее наибольшее закодированное значение из диапазона этой категории.

Пример

Код продукции нефтепереработки. В этом коде каждая продукция представляется четырехзначным числом, взятым последовательно из рядов, характеризующих группировку продукции.

Такими группами продукции, которые были определены заранее, и соответствующими диапазонами закодированных значений являются:

- ***газ*** ***1000—2999***
- ***бензин и топливо*** ***3000—4999***

Преимущества:

- быстрое присвоение закодированных значений;
- компактность;
- простота подтверждения допустимости закодированных представлений.

Недостатки:

- максимальная вместимость используется не полностью.

Примечание — Эту систему не следует путать с иерархическими кодами (см. 6.7). Групповое последовательное кодирование следует использовать только в тех случаях, когда категории стабильны и нет вероятности того, что элемент набора может относиться к разным категориям (в текущий момент или в обозримом будущем).

6.2.3.4 Упорядоченные последовательные коды

Эти коды не являются абсолютно последовательными кодами, хотя закодированные значения могут казаться такими.

Этот тип кодов может быть успешно использован, только если все элементы набора заранее известны и набор не может расширяться.

До присвоения закодированных значений элементы набора упорядочиваются на основании некоторой характеристики, например, в алфавитном порядке имен или хронологически (для событий, действий). Полученный таким образом порядок выражается с помощью закодированных значений, которые сами последовательно выбираются из упорядоченного списка.

Такая схема кодирования делает возможным легкое получение и сортировку экземпляров элементов. Это используется в тех случаях, когда невозможно или трудно получить требуемые результаты с помощью экземпляров незакодированных элементов данных, например, из-за присутствия разделяющих слова пробелов или в результате цифрового выравнивания.

Преимущества:

- быстрое присвоение закодированных значений;
- компактность;
- простота подтверждения допустимости закодированных представлений.

Недостатки:

- максимальная вместимость используется не полностью.

Пример — Во Франции «Департаменты» были закодированы числами на основании списка, отсортированного в алфавитной последовательности. Эти коды могли быть элементами других кодов, например, почтового кода, национального идентификационного номера, регистрационного номера транспортного средства.

Недостатком этой системы является то, что она не может быть приспособлена к возможным последующим изменениям. Рассмотрим пример: «Departement Corse» было присвоено кодовое значение 20. Когда «Corse» был разделен на «Haute Corse» и «Corse du Sud» пришлось использовать буквы (2A и 2B соответственно) в цифровом коде, что создало аномалии.

6.2.4 Заключение

При выборе последовательного кода, код абсолютно возрастающего типа представляет собой наилучшее решение по гибкости, затратам и простоте присваивания.

6.3 Случайные коды

6.3.1 Принципы

Элементам набора присваиваются закодированные значения из набора возможных неупорядоченных значений или с помощью некоторого алгоритма. Нет корреляции между элементами набора и их закодированными значениями. Этот метод отличается от последовательного кодирования тем, что в последнем для присвоения закодированных значений используются последовательно упорядоченные диапазоны.

Цель данного метода состоит в том, чтобы сделать невозможной какую-либо интерпретацию кода.

6.3.2 Преимущества:

- простое и быстрое присвоение закодированных значений, возможно — автоматическое;
- компактность;
- использование максимальной вместимости.

6.3.3 Недостатки:

- с помощью закодированных представлений невозможна классификация или группировка элементов набора;
- необходим предварительно установленный список или алгоритм создания случайных чисел, исключающий любое дублирование чисел.

6.3.4 Использование

Случайные коды могут быть использованы как самодостаточные коды для целей идентификации или как часть составного кода, другие части которого основаны на других принципах кодирования.

6.4 Мнемонические коды

6.4.1 Принципы

Элементам набора присваивают закодированные представления, беря один или несколько символов из их названий.

Цель этого метода состоит в том, чтобы повысить эффективность запоминания пользователями закодированных значений.

6.4.2 Преимущества:

- простота запоминания пользователями закодированных значений, что избавляет от частого обращения к спискам кодов;
- облегчение выявления ошибок.

6.4.3 Недостатки:

Кодирование существенно зависит от способа исходного выражения элементов набора (языка, единиц измерения и проч.). Существует риск, что созданные мнемоники в языке пользователя имеют смысл, не связанный с представляемыми экземплярами набора.

6.4.4 Использование

Мнемонические коды могут успешно применяться для ограниченных идентифицирующих кодовых наборов, которые достаточно стабильны и названия элементов которых хорошо известны в среде пользователей.

Пример 1

По ИСО 3166 — коды для представления названий стран:

Название категории	*Код страны*
Австрия	*AT*
Канада	*CA*
Франция	*FR*
США	*US*

Пример 2

По Рекомендации 20 Экономической комиссии ООН для Европы (UN/ECE) — коды единиц измерения

Название единицы измерения	*Закодированное представление*
Сантилитр	*CLT*
Сантиметр	*CMT*
Кубический сантиметр	*CMQ*
Кубический фут	*FTQ*
Градус Цельсия	*CEL*
Градус Фаренгейта	*FAH*
Галлон	*GLI*
Килограмм	*KGM*
Километр	*KMT*
Киловатт	*KWT*
Метр	*MTR*
Минута	*MIN*
Штука	*PCE*
Квадратный сантиметр	*CMK*
Квадратный фут	*FTK*

6.5 Коды, основанные на сокращениях

6.5.1 Принципы

Существенной характеристикой этого кода является метод сокращения обозначений элементов набора, который изначально единообразно определен.

Цель этого метода состоит в создании кратких представлений непосредственно из исходных названий элементов набора. Этот метод похож по своим целям на мнемонический код.

6.5.2 Преимущества

Простое сжатие иногда весьма объемных данных.

6.5.3 Недостатки

Уникальность получающихся в результате процесса сокращения закодированных значений не может быть гарантирована без проверки всех закодированных значений после добавления каждого нового.

6.5.4 Использование

Этот тип кода рекомендуется использовать для больших наборов элементов, когда требуется уменьшить размеры представлений элементов.

Пример

Коды ответов телекса.

Код ответа является строкой символов, которая автоматически передается телекс-терминалом после получения им управляющего символа WRU («кто ты?»). По соглашению длина этой строки ограничена 14 символами.

Эти закодированные представления содержат переменное, но не превышающее 14, число символов. Наиболее часто коды создаются путем сжатия названия организации. Закодированные представления не обязательно должны включать в себя номер телекса.

Организация	*Номер телекса*	*Код ответа*
Comptoir des viandes europeennes 5, rue de la Corderie, RUNGIS	*204 045*	*CEV 204 045F*
Tuilerie generale du midi RN de Clermont, BEDARIEUX	*490 695*	*TEGEMIDI BEDAR*
BENDIX Automobile Division de DBA	*611 138*	*DBABX 611138*

6.6 Матричные коды

6.6.1 Принципы

Матричный код основывается на наличии таблицы с двумя входами. Значения, присвоенные строкам и графам этой таблицы, используются для создания закодированного представления элементов набора в данной таблице на соответствующих координатах.

Цель данного метода состоит в том, чтобы присвоить осмысленные закодированные значения элементам набора, которые имеют ряд общих характеристик в различных комбинациях.

6.6.2 Преимущества:

- интерпретируемость закодированных значений;
- простота присваивания закодированных значений.

6.6.3 Недостатки:

- необходимость предварительно введенных таблиц, охватывающих все характеристики элементов кодируемого набора;
- трудность приспосабливания к новым требованиям, таким как новые или измененные характеристики, или новые комбинации.

6.6.4 Использование

Матричные коды могут быть полезными для целей идентификации элементов набора, когда их характеристики хорошо структурированы и стабильны.

Пример

Следующая таблица может быть использована для кодирования электрических ламп на основании их характеристик. Например, кодовое представление 04 B01 стандартной лампы на 220/230 вольт мощностью 60 ватт с патроном B22 может быть интерпретировано следующим образом:

04 — номер таблицы;

B — номер строки, соответствующей мощности 60 ватт;

01 — номер графы, соответствующей патрону B22.

Таблица 04		*Стандартная лампа 220/230 В*			*Стандартная лампа 250/260 В*	
		Патрон			*Патрон*	
		01	*02*	*03*	*04*	*05*
40 Вт	*A*	*B22*	*E27*	*B22*	*E27*	
60 Вт	*B*	*B22*	*E27*	*B22*	*E27*	
75 Вт	*C*	*B22*	*E27*	*B22*		
100 Вт	*D*	*B22*	*E27*	*B22*	*E27*	
150 Вт	*E*	*B22*	*E27*	*E27*		
200 Вт	*F*	*B22*	*E27*	*E10*		

6.7 Иерархические коды

6.7.1 Принципы

Иерархический код основан на последовательном и нарастающем разбиении кодируемого набора элементов на маленькие группы.

Каждая группа более высокого иерархического уровня включает в себя все группы своих более низких уровней и только эти группы. Этот тип кода основывается на различиях между характеристиками элементов на каждом уровне.

Характеристики каждого уровня должны быть взаимоисключающими.

6.7.2 Преимущества:

- простота классификации или группировки элементов;
- возможность отчетов на верхних уровнях агрегации;
- интерпретируемость закодированных значений.

6.7.3 Недостатки:

- ограниченное использование теоретической мощности;
- потеря гибкости из-за сложности использованного принципа (когда уровни могут стать переполненными или высшие уровни нельзя будет переупорядочить без полного пересмотра кода);
- необходимость следовать по кодам уровней от высших к низшим с целью присвоения или интерпретации кода;
- сложность, зависимость от числа уровней, риск повторного введения уже использованных на верхних уровнях характеристик.

6.7.4 Использование

В общем случае иерархические коды используют в целях классификации. Количество уровней устанавливают в зависимости от информационных требований. Иерархические коды менее пригодны для целей идентификации и указания.

Иерархические коды весьма подходят, например, для статистических задач, для отчетов о движении продукции, для классификации публикаций по темам. На практике встречаются как фиксированные, так и переменные форматы. Фиксированные форматы более просты для обработки, чем переменные.

Пример 1

Фиксированный нарастающий формат

Из Гармонизированной системы описания и кодирования товаров

Формат кода гармонизированной системы состоит из 6 числовых позиций, разделенных на независимые сегменты по 2 позиции в каждом.

37	***Фотографическая и кинематографическая продукция***
3705	***Экспонированные и обработанные фотографические пластинки и пленки, отличные от кинематографических***
370520	***Микрофильмы***

Пример 2

Переменный нарастающий формат

Из Универсальной десятичной классификации (УДК).

Количество символов и сегментов в закодированных представлениях переменно, так что степень подробности описания может быть продолжена до нужного уровня. Понятие «скаты крыш в архитектуре» выражается закодированным представлением 624.024.13:

624	***Гражданская промышленность***
624.02	***Конструктивные компоненты***
624.024	***Крыши, кровли***
624.024.13	***Скаты крыши***

6.8 Сопоставляющие коды

6.8.1 Принципы

Сопоставляющие коды являются составными кодами, состоящими из сегментов, которые предоставляют характеристики, выбранные для представления элементов набора. Эти характеристики не зависят друг от друга. Комбинируемые таким образом закодированные представления могут быть любых типов (последовательными, мнемоническими, случайными).

6.8.2 Преимущества:

- простота группировки элементов набора на основании одной или нескольких характеристик, выраженных в закодированных значениях;
- мощность, связанная с количеством значений, которые может принимать каждая характеристика;

- интерпретируемость закодированных значений.

6.8.3 Недостатки:

- требование включать большое количество характеристик может привести к большому числу символов в каждом закодированном значении;
- трудно приспособить к требованиям для новых характеристик.

6.8.4 Использование

Сопоставляющие коды очень подходят для классификации предметов, имеющих несколько общих характеристик. Областями применения являются, например, трассирующее кодирование (какая продукция, когда и где произведена) или групповые технологические подходы к разработке и производству.

Пример

Кодирование металлов:

NNNN	***NN***	***NNN***
марка	***форма***	***размеры***

Эти три типа характеристик (марка, форма и размеры) в достаточно большой степени не зависят друг от друга.

6.9 Комбинированные коды

6.9.1 Принципы

Комбинированные коды являются составными кодами, которые собраны из сегментов, предоставляющих разные характеристики элементов набора. Эти характеристики зависят друг от друга и иерархически взаимосвязаны. В этом отношении комбинированные коды отличаются от сопоставляющих.

6.9.2 Преимущества:

- простота присвоения закодированных значений;
- удобство распределения и сопровождения закодированных значений;
- возможность некоторого расширения интерпретируемости закодированных значений;
- удобство проверки допустимости закодированных значений.

6.9.3 Недостатки:

- теоретическая мощность не может быть использована в полной мере.

6.9.4 Использование

Комбинированные коды часто применяются в целях идентификации, охватывая широкую область приложений.

Пример из Европейского цифрового кода продукции (EAN)

Код EAN				
nnnnnnnnnnnn				***Полная идентификация EAN***
nn				***Страна***
	nnnnn			***Производитель***
		nnnnn		***Идентификация продукции***
			n	***Контрольная сумма***

6.10 Коды, суммирующие значения

6.10.1 Принципы

С этим типом кода для представления элементов набора допустимы только числа, являющиеся степенями 2. Код такого типа может применяться, когда элементы набора должны комбинироваться.

Закодированными представлениями экземпляров являются степени 2:

1, 2, 4, 8, 16, 32, 64, 128, 256, 512 и т. д.

Таблица значений получается путем присвоения значения степени 2 каждому экземпляру:

Экземпляр	Значение	Экземпляр	Значение
1-й	1	6-й	32
2-й	2	7-й	64
3-й	4	8-й	128
4-й	8	9-й	256
5-й	16	10-й	512

6.10.2 Преимущества

Компактность.

6.10.3 Недостатки

Необходим алгоритм для присваивания и интерпретации.

6.10.4 Использование

Этот метод подходит для небольших наборов элементов, требующих много комбинаций при условии, что их порядок несущественен. Используя таблицу, можно присвоить уникальное закодированное представление любой комбинации, суммируя все имеющиеся элементы.

Пример

Комбинация элементов 1, 8, 4 и 9 будет закодирована как 393, что является суммой 1, 128, 8 и 256.

Зная таблицу атрибутов кода этого типа, всегда можно идентифицировать элементы, так как при данном выборе чисел (степени 2) 393 может быть только суммой 256+128+8+1.

7 Характеристики кодов

7.1 Общие характеристики

Общими характеристиками кодов являются:

- однозначность;
- расширяемость;
- компактность;
- простота;
- многосторонность;
- пригодность для сортировки;
- стабильность;
- выразительность;
- размер;
- структура и формат;
- мощность.

7.2 Однозначность

Однозначность является основной характеристикой любого кода.

Однозначность достигается, когда одному и только одному элементу набора соответствует один закодированный элемент, и наоборот.

7.3 Расширяемость

Это свойство означает способность кода приспосабливаться к требованиям добавления новых элементов в нужных местах кодируемого набора без перекодирования или реструктуризации всего кода. Код должен допускать расширение существующих делений и добавление новых.

Планируемые сроки использования кода должны учитывать количество элементов набора, которые должны быть закодированы, и стабильность или динамику окружения, управляющего расширением набора элементов.

7.4 Компактность

Число символов в закодированном представлении регулирует компактность кода.

Закодированное представление должно иметь наименьшее количество символов, необходимое для кодирования текущих и будущих элементов набора.

Краткое закодированное представление дает преимущества при:

- вводе данных;
- обмене данными электронным или иными способами;
- хранении в базе данных;
- обработке данных.

7.5 Простота

Метод кодирования должен быть прост для понимания и применения пользователями. Компактность и выбор символьного набора при этом не менее важны, чем выбор понятий, определение структур, отсутствие исключений и отклонений, эффективность процедур присваивания.

7.6 Многосторонность

Код должен быть применим в различных взаимосвязанных прикладных областях. Это свойство обеспечивает интеграцию систем.

7.7 Пригодность для сортировки

Если есть требование создавать или обрабатывать данные в предписанном порядке, то код должен удовлетворять этому требованию.

Пригодность для сортировки существенно зависит от типа выбранного кода.

В составных кодах (например, иерархических) могут быть выбраны более или менее сложные возможности сортировки и/или группировки в зависимости от характера элементов набора и назначения системы. Эти возможности зависят в основном от количества сегментов, предусмотренных в структуре кода, или выразительности закодированных значений.

7.8 Стабильность

Код стабилен, когда он допускает запланированные изменения без модификации всей структуры кода. На уровне закодированных представлений эта концепция выражается в терминах стабильности представляемых уникальных элементов.

Пользователям нужны стабильные коды. Присваивание закодированных значений должно проводиться с наименьшей возможной вероятностью модификаций как самих закодированных значений, так и структуры кода.

Модификации дорогостоящи, приводят к росту количества ошибок, требуют времени и ресурсов и могут повредить всю систему, если ими не управлять должным образом.

Закодированные представления, которые удаляются, когда элементы выпадают из набора, не должны использоваться повторно для других элементов.

7.9 Выразительность

Код называют выразительным, если кодовые представления прямо (например, мнемонические коды) или косвенно через ссылку на одну или несколько таблиц (например, иерархические или сопоставляющие коды) выражают заложенный в них смысл.

Выразительность так же связана с возможностями сортировки и группировки на основании характеристик элементов набора, используя их закодированные представления.

На практике выразительность важна в случае классификации. Для задач идентификации и указания рекомендуются невыразительные коды.

7.10 Размер

Размер выражается как число позиций закодированного представления.

Закодированные значения могут быть определены с фиксированным или переменным количеством символов.

Закодированные значения с переменным количеством символов имеют два больших недостатка:

1 Количество символов непредсказуемо, что может привести к проблемам выравнивания при хранении закодированного значения в поле данных, которое содержит больше символов, чем использовано для закодированного значения.

Данные могут быть выровненными вправо, влево или на десятичную точку.

Ошибки левого и правого выравниваний являются обычными в записях данных (сдвиг поля). Некорректное использование закодированных представлений, имеющих переменное количество символов, в полях данных фиксированного формата может привести к разным представлениям (например, **12, *12* или 12**, где * представляет символ SPACE), которые не будут распознаны устройствами обработки как одно и то же закодированное представление.

2 Ошибки добавления или пропуска символов не могут быть просто выявлены человеком или машиной.

Неконтролируемое добавление суффиксов и/или префиксов к исходному закодированному значению является общей и часто встречающейся проблемой. Так как суффиксы и префиксы обычно имеют различный характер и не всегда должны присутствовать, то это легко может привести к увеличению неправильных интерпретаций и ошибок.

По этим причинам рекомендуется использовать закодированные значения с фиксированным количеством символов.

7.11 Структура и формат

Формат (или структура символов) закодированного значения предпочтителен полностью цифровой или полностью алфавитный. Смешанные форматы (алфавитно-цифровые) могут использоваться, когда символы в конкретных позициях, например первый или последний символы, всегда цифровые или алфавитные. Случайные алфавитно-цифровые форматы использовать не следует, так как при этом имеется существенный риск человеческих ошибок.

Структура кода определяет:

- количество позиций или групп позиций, образующих закодированное представление;
- набор допустимых символов для каждой позиции.

Примечание — Составной частью структуры могут быть пробелы.

Входной контроль для выявления синтаксических ошибок относится, главным образом, к структуре. Таким образом, каждая позиция закодированного представления каждой группы из нескольких позиций должна быть определена как алфавитная, цифровая, алфавитно-цифровая или специальная.

7.12 Мощность

Эта характеристика есть количество закодированных представлений, которые могут быть сформированы комбинацией используемых символов для каждой позиции, учитывая выбранное основание счисления.

Примеры

1 Для 1 позиции и основания 2 с двоичными символами мощность C=2.

2 Для 1 позиции и основания 10 с десятичными цифровыми символами C=10.

3 Для 1 позиции и основания 26 с алфавитными символами C=26.

Это теоретические мощности. Они предполагают использование всех комбинаций всех символов.

Ограничения, вызванные практическими или концептуальными причинами, уменьшают эти теоретические мощности.

Для вычисления мощности данного кода для охвата всех ситуаций при сохранении его однозначности используют следующую формулу (принимая, что используется 24 алфавитных символа и 10 цифр, так как букв I и O следует, по возможности, избегать):

C = (24A) + (10N),

где C — общее число доступных возможных комбинаций кода,

A — число алфавитных позиций в коде,

N — число цифровых позиций в коде,

(A+N для комбинированного кода равно общему числу позиций в коде).

Примечание — В приведенной выше формуле принято, что данная позиция является либо алфавитной, либо цифровой. Если одна и та же позиция может быть как алфавитной, так и цифровой, то формула принимает следующий вид:

C = (36)(A+N) или

C = (34)(A+N), когда буквы I и O не используются.

Позиции, занятые контрольными символами, не должны учитываться при вычислении мощности.

На практике выбор мощности является компромиссом между:

1) предполагаемым расширением системы,

2) ограничениями на число символов, составляющих закодированное представление,

3) простотой написания и использования закодированных представлений,

4) желательным полезным функционированием системы,

5) стоимостью работы и т. д.

Мощность следует рассматривать как предел: она отличается от числа закодированных представлений, существующих в коде в любой заданный момент времени, и от числа закодированных представлений, которые были присвоены с момента введения кода (некоторые из них со временем могли быть удалены).

8 Представление закодированных значений

8.1 Наборы символов

Свойством любого кода является то, что он есть метод представления закодированных значений, при котором используются различные символы.

Код называют:

а) цифровым, если в нем используются только цифры,

б) алфавитным, если в нем используются только буквы,

в) алфавитно-цифровым, если в нем используются как цифры, так и буквы, возможно, вместе со специальными символами.

Цифры

В качестве цифр используют арабскую нумерацию: 0, 1, 2, 3, 4, 5, 6, 7, 8 и 9.

Буквы

В качестве алфавита используют латинский.

В зависимости от требований можно использовать весь алфавит из 26 букв или его часть. Использование национальных алфавитов в закодированных представлениях находится вне области применения настоящего стандарта.

Рекомендуется использовать буквы только одного регистра (строчные или прописные).

Специальные символы

Специальные символы можно использовать в закодированных представлениях в дополнение к алфавитным (&, $, ...); в таком случае мощность возрастает, и становится доступным семейство символов, которые могут быть зарезервированы для специальной обработки.

Следует избегать следующих символов, которые:

1) не относятся к инвариантному набору ИСО 646 IRV;

2) могут быть ошибочно интерпретированы или неправильно переписаны. Так, по мере возможности, следует избегать пробелов.

Пример — 123 ABC следует записывать как 123ABC, так как пробел не имеет смысловой нагрузки и может быть опущен при переписывании.

3) могут использоваться в качестве служебных в синтаксических схемах для обмена данными.

Пример — знак вопроса (?), двоеточие (:), знак плюс (+), апостроф (') используют в качестве служебных символов в классе А ИСО 9735 (EDIFACT). Однако для них предусмотрены управляющие последовательности.

Примечание — Допускается использование альтернативного набора символов путем расширения символьных наборов, которые зарегистрированы в соответствии с ИСО 2375. Если специфицирован набор символов, отличный от ИСО 646 IRV, то издающей организацией должно быть идентифицировано подмножество набора символов, которое должно использоваться для закодированных представлений.

8.2 Способы представления

Имеется несколько вариантов физического представления закодированных значений. Какой из них будет выбран, зависит от прикладной области и имеющихся устройств.

Для ручной обработки предпочтение обычно отдается удобочитаемым для человека закодированным представлениям. В этом случае закодированные значения будут представляться традиционным способом с помощью символов латинского алфавита и арабской нумерации. Такое представление будет также выбрано для компьютерного вывода, при печати бумажных документов или листингов и при выводе на устройство визуального отображения.

Если требования включают в себя обработку механическими или электронными средствами, то доступны другие методы. Одним из наиболее часто используемых методов является автоматическая идентификация, для которой широко используется метод штрих-кодов. Существует ряд символов для штрих-кодов. Выбор символов должен проводиться достаточно тщательно. Он зависит от прикладной области, групп пользователей и принятых в среде функционирования соглашений.

Для числовых закодированных значений в общем случае применяют два метода:

- Символика, используемая в розничной торговле для указания цены, UPC (Universal Product Code — универсальный код продуктов) для Северной Америки и EAN (European Article Number — Европейский номер продукта) для остального мира. Эта символика используется, в основном, для идентификационных номеров продукции на потребительских товарах.
- Чередующиеся 2 из 5.

Эта символика кодирует цифровые данные в формате высокой плотности.

Для алфавитно-цифровых закодированных значений наиболее популярной символикой является:

- Код 3 из 9 (код 39).

Эта символика кодирует весь набор, состоящий из прописных букв, цифр и некоторых других символов, в шрих-код переменной длины.

Кроме штрихового кодирования существуют другие методы автоматической идентификации, такие как оптическое чтение символов OCR или магнитные полосы. Другим примером разработок являются смарт-карты, позволяющие хранить большой объем данных, которые могут быть обработаны.

9 Проектирование кодов

Настоящий раздел предназначен для поддержки проектирования кодов. Это может помочь избежать потенциально обширных последовательностей неадекватно задуманных и разработанных кодов.

Требования часто могут противоречить друг другу. Например, если структура кодирования должна быть достаточно расширяемой для будущих потребностей, то можно в некоторой степени пожертвовать компактностью. Следовательно, для получения оптимальной эффективности в данной прикладной области нужно тщательно рассматривать все аспекты и проводить соответствующие оценки.

9.1 Общие положения

9.1.1 Подготовка

Когда требуется новый код, необходимо затратить достаточно времени и усилий для предварительного изучения, определения и планирования деятельности. Следует предвидеть потенциальные проблемы и тщательно оценить все альтернативы до внедрения нового кода.

9.1.2 Использование существующих кодов

По мере возможности, следует использовать существующие коды. Новые коды должны разрабатываться только тогда, когда это абсолютно необходимо. Всегда следует учитывать предпочтения пользователей. Будет правильно рассмотреть все схемы кодирования, используемые предполагаемыми пользователями нового кода.

Предпочтение следует отдавать принятию одобренных международных кодов, которые применяются в стандартах ИСО, рекомендациях и справочниках Экономической Еврокомиссии ООН и других международных организаций.

9.1.3 Выразительность кода

При надлежащем использовании выразительные коды дают основу для дополнительной информации и имеют тенденцию быть проще и более читабельными для людей, чем невыразительные коды. Однако, следует проявлять осторожность при разработке выразительных кодов в том отношении, что выразительные части должны быть связаны со стабильными категориями. Например, выразительный

код организации не должен быть связан с ее расположением, так как изменение последнего приведет к изменению кода. Чрезмерно выразительные коды могут стать неуправляемыми и потерять расширяемость, поэтому их следует избегать.

Для большинства задач идентификации и для всех задач указания рекомендуется использовать невыразительные коды. Преимущества и недостатки различных методов кодирования приведены в разделе 6.

9.1.4 Кратная совместимость набора кодов

В некоторых случаях для удовлетворения требований системы требуется несколько кодов или представлений. Единственный код является идеальной целью, но не всегда — наиболее практичным решением. Если необходимо несколько кодов, то они должны допускать преобразования из одного кода в другой, т.е. элементы остаются неизменными, а коды различаются.

9.1.5 Мнемонические коды

Мнемонические коды могут использоваться с целью ассоциаций и запоминания, повышая тем самым эффективность обработки человеком, при условии, что они не используются для идентификации очень длинных, нестабильных списков элементов. Однако, следует весьма тщательно выбирать мнемонические структуры для обеспечения надлежащей гибкости. Следует избегать мнемоники, если закодированный набор потенциально является очень большим, так как ее эффективность падает с ростом числа кодируемых элементов. Когда в системе мнемонические или другие осмысленные закодированные представления не могут быть обеспечены для всех элементов набора, следует отдавать предпочтение элементам, имеющим наибольшую частоту использования.

9.1.6 Наименование кода

Все независимые сегменты кода должны быть индивидуально поименованы с помощью стандартно, однозначно и согласовано используемых меток.

9.1.7 Вычисление мощности кода

При вычисления мощности данного кода для охвата всех ситуаций, в которых сохраняется его однозначность, используют формулы, приведенные в 7.12.

9.2 Количество символов

9.2.1 Компактность

Для экономии места и уменьшения времени передачи данных закодированные значения должны состоять из минимального количества символов, но в то же самое время, должны быть оптимизированы относительно возможностей пользователей кода.

9.2.2 Фиксированное количество символов

Коды с фиксированным количеством символов (например, всегда три символа, а не один, два или три) более надежны и легче в использовании, чем коды переменной длины.

9.2.3 Сегментация

Закодированные значения, имеющие более четырех алфавитных или пяти цифровых символов следует делить на более мелкие сегменты с целью более надежной записи, например, XXX-XXX-XXXX более надежно, чем XXXXXXXXXX.

9.2.4 Возможное расширение

Структура кода должна обеспечивать добавление к набору новых элементов без перекодирования существующих или расширения формата закодированных представлений.

9.3 Формат закодированных представлений

9.3.1 Требования пользователей

Компоненты и сегменты кода следует форматировать в соответствии с информационными потребностями пользователей, учитывая максимальную простоту просмотра кода на предмет точности и полноты и компактность содержимого данных.

9.3.2 Алфавитные и цифровые коды

В общем случае человеку легче читать цифровые, чем алфавитные закодированные значения, если в последних не используется мнемоника. Контролируемые алфавитно-цифровые закодированные значения (т. е., те, в которых конкретная позиция всегда является цифровой или алфавитной) более удобны для чтения, чем случайные. Например, формат AA999 (первые два символа всегда буквы, а последние три — цифры) легче для чтения, чем формат, в котором буквы и цифры появляются в произвольных позициях.

9.3.3 Группирование символов

В случае, когда код построен из алфавитных и цифровых символов, аналогичные типы символов следует группировать, а не разбрасывать по всему закодированному представлению. Например, меньше ошибок будет встречаться в трехсимвольном коде со структурой буква-буква-цифра (HW5), чем в последовательности буква-цифра-буква (H5W).

9.3.4 Последовательность позиций кода

Если код делит полный набор категорий на небольшие группы, то позиции высших порядков должны соответствовать более широким, общим категориям; позиции низших категорий должны быть более селективными (включая суффиксы). Примером является числовое представление даты в соответствии с ИСО 8601 (YYMMDD). Если составной код состоит из нескольких независимых кодов, то выбор кодового сегмента, занимающего позиции более высокого порядка, должен основываться на рассмотрении требований использования и эффективности обработки.

9.3.5 Разделение сегментов кода

Сегменты кода следует разделять дефисом (при отображении) или располагать совершенно отдельно друг от друга (при хранении и отображении), если позиции или сегменты совершенно независимы друг от друга и могут использоваться по отдельности (т.е. для понимания их смысла не требуются другие части кода).

9.3.6 Контрольные символы

Следует рассмотреть добавление выявляющего ошибки символа во избежание этих ошибок при записи. Использование самопроверяющегося кода предотвращает многие необязательные проблемы передачи данных, связанные с неправильной записью и предоставлением ошибочной информации.

9.4 Наборы символов

9.4.1 Специальные символы

В структуре кода следует использовать обычные символы и избегать символов, отличных от букв и цифр таких, как дефис, точка, пробел, звездочка и т.д. (за исключением разделителей сегментов кода, когда могут быть использованы дефис или пробел). В коде должны использоваться буквы только одного регистра, т. е. ABC...Z или abc...z. В именах, названиях и сокращениях могут использоваться буквы обоих регистров и другие символы. Словарь для данной системы кодирования должен содержать как можно меньше классов символов. По мере возможности, используемый набор символов должен соответствовать 7-битовому кодовому набору для обмена информацией (см. ИСО 646).

9.4.2 Визуальное сходство

Когда необходимо использовать алфавитно-цифровою случайную структуру кода, следует избегать символов, которые легко перепутать с другими. Вот некоторые примеры: буква I и цифра 1, буква O и цифра нуль, буква Z и цифра 2, буква G и цифра 6, буквы B и S и цифра 8, буквы O и Q.

9.4.3 Акустическое сходство

В невыразительных кодах следует избегать символов, которые могут быть перепутаны при произношении (акустически однородны); например, буквы B, D, G, P и T или буквы M и N.

9.4.4 Гласные

Следует избегать использования гласных (A, E, I, O и U) в алфавитных кодах или в позициях кодов, имеющих три и более последовательных алфавитных символа, для исключения неумышленного образования слов обычного языка.

9.4.5 Учет сортировки

В любой конкретной позиции символ должен быть либо алфавитным, либо цифровым во избежание несовместимости с сортировкой последовательности.

9.5 Соглашения о присваивании

9.5.1 Правила кодирования

Должны быть четко установлены правила кодирования, и эти правила должны последовательно применяться. Например, мнемонические сокращения могут образовываться путем удаления всех гласных из названий кодируемых элементов как в случае DT для date (дата) или GRN для green (зеленый), или могут использоваться первые буквы слов кодируемых элементов как в случае EOF для End of File (конец файла).

9.5.2 Коды для количественных данных

Не следует кодировать количественные данные или денежные суммы, так как это приводит к дополнительным преобразованиям и потере точности. Например, можно было бы закодировать числа от 1 до 99 как A, 100—199 как B и т. д. Такое кодирование может оказаться желательным для задач категоризации, но статистическая значимость будет потерена, так как из закодированных значений не могут быть получены фактические числа. Категоризация может быть осуществлена при последующей обработке, а не при кодировании входных данных.

9.5.3 Использование «естественных» данных

Не следует разрабатывать кодовую структуру, если конкретные данные в их естественной форме (такие, как процентное содержание) являются вполне пригодными и адекватными.

9.5.4 Использование «0000» и «9999» в качестве закодированных значений

Не следует использовать последовательности из «0» или «9» в качестве присвоенных закодированных значений. Эти значения следует зарезервировать для специальных ситуаций или для индикаторов обработки.

9.5.5 Категория «разное»

С большой осторожностью нужно использовать кодирование категорий «разное» или «прочее». Нельзя допускать размещение в этих категориях объектов, которые фактически относятся к более конкретному классу.

9.6 Выявление ошибок с помощью контрольных символов

Примечание — Подробное рассмотрение контрольных символов приведено в ИСО 7064.

Некоторые ошибки, возникающие в результате участия человека в преобразованиях закодированных представлений, могут быть выявлены с помощью контрольных символов. Они получаются путем применения арифметических операций к символам, образующим закодированное представление.

Операция взятия модуля (деления) дает основу для присвоения каждому закодированному представлению контрольного символа.

Метод вычисления контрольных символов определяет эффективность возможности выявления ошибок. Могут быть выявлены и, иногда, исправлены ошибки, включающие в себя:

1) единственный символ: 12545 вместо 12345;

2) два символа: 22335 вместо 12345;

3) перестановку символов: 21345 вместо 12345.

ИСО 7964 устанавливает системы контрольных символов. Для удовлетворения потребностей различных приложений может быть выбран только небольшой набор совместимых систем. Эти системы обеспечивают возможность достижения наибольшей степени защиты от типичных ошибок при ограничениях каждого приложения.

9.7 Рекомендации для некоторых прикладных областей

9.7.1 Классифицирующие коды для логистики и управления

При планировании и подготовке отчетов не используют идентифицирующие коды. Для рассматриваемых целей они слишком подробны. Элементы должны быть сгруппированы на основании определенных критериев. Для долгосрочного планирования обычно более приемлемы данные по группам на высших уровнях агрегации, а не данные по отдельным продуктам.

Когда нужно провести агрегацию на различных уровнях, то в общем случае наилучшим решением является иерархический код.

9.7.2 Классифицирующие коды для разработки и производства

Широкое распространение получили реализации систем CAD (Computer Aided Design — автоматизированное проектирование) и CAM (Computer Aided Manufacturing — автоматизированное производство). Для интеграции этих систем требуются кодирование и классификация. Важным приложением является групповая технология.

В групповой технологии классификация используется для определения сходства видов продукции или производственных процессов. Кодирование в групповой технологии направлено на присвоение осмысленных идентифицирующих символов взятым по отдельности или в логической группировке внутренним характеристикам классифицируемых объектов и процессов.

Для приложений такого типа наиболее подходят сопоставляющие коды.

9.8 Практические рекомендации

Настоящие практические рекомендации даны на основе преимуществ и недостатков различных типов кодов.

9.8.1 Выбор структуры

Последовательные цифровые коды (см. 6.2) являются более предпочтительными по сравнению с кодами других типов по следующим причинам:

- в них максимально используется теоретическая мощность;
- арабская нумерация используется в большинстве стран мира;
- при ручном наборе или запросе использование цифровых клавиш проще и быстрее;
- при устной передаче они проще и надежней;
- они облегчают использование контролирующих символов.

Мнемонические алфавитные коды (см. 6.4 и 6.5) следует зарезервировать преимущественно для коротких кодов, когда вопросы запоминания рассматриваются как приоритетные.

К недостаткам мнемонических алфавитных кодов относится следующее:

- существенная зависимость от используемого языка;

Пример — В ИСО 3166 алфавитный двухбуквенный код построен из букв того названия страны, которое в этой стране используется; таким образом, этот код не дает мнемонических преимуществ для пользователей, которые не знают соответствующего языка. Например, «DZ» (Djazair) для Алжира «DE» (Deutschland) для Германии.

- необходимость изменяться при изменении названия элемента;

Пример — При изменении названия страны Верхняя Вольта на Буркина-Фасо пришлось менять код с «HV» (Haute Volta) на «BF».

- практическая недостижимость теоретической мощности, так как настоятельно не рекомендуется использовать буквы, приводящие к путанице при произношении (B, D, G и т.д.) или написании (O и Q).

9.8.2 Выбор длины

Количество символов в коде непосредственно зависит от трех факторов:

- выбранной структуры;
- требуемой мощности;
- желательного фактора расширения.

Код фиксированной длины использовать проще и надежнее, чем код переменной длины.

9.8.3 Мощность

Часто реальная мощность алфавитного или алфавитно-цифрового кода намного меньше его теоретической мощности, если принять во внимание, что некоторые символы и их комбинации не используются во избежание путаницы.

Пример — Ранние системы телефонных номеров, например, в Париже, были построены на 3 буквах и 4 цифрах. В алфавитной части, являвшейся мнемоническим сокращением названия телефонной станции, многие комбинации не использовались. Замена структуры кода на 7 цифровых символов позволила увеличить истинную мощность и облегчила ее последующее расширение до 8 цифр.

9.8.4 Потенциальное расширение

При выборе длины кода следует принять во внимание пригодность полученной мощности для последующих расширений.

Пример — При внедрении системы кодирования для набора из 90 текущих элементов (например, районов страны) и разработке двухзначного цифрового кода имеется большая вероятность того, что через какое-то время мощность окажется недостаточной.

9.8.5 Составные коды

Для этих кодов следует избегать использования сегментов (см. 6.8 и 6.9), которые сопровождаются разными организациями. Каждое изменение, принятое другой сопровождающей организацией, вызовет изменения в используемом составном коде.

9.8.6 Сопровождение и управление версиями кодовых наборов

Изменения и (или) расширения кодовых наборов могут быть такими, что повлияют на существующие приложения. Следует определить правила, соответствующие правилам сопровождения и управления версиями для данного типа элементов данных, которые гарантируют недвусмысленность при использовании разных версий кодовых наборов. Существенные изменения семантики элементов за-

кодированного набора должны приводить к новому закодированному набору, узнаваемому, например, по номеру версии или дате.

Пример — Международные сокращения терминов доставки (INCOTERMS) со времени своей первой публикации в 1953 г. существенно изменялись и расширялись в 1980 и 1990 гг. Изменения и расширения привели к отдельным закодированным наборам, которые уточняются датой, т.е. INCOTERMS 1953, INCOTERMS 1980, INCOTERMS 1990.

Требуются также правила повторного использования устаревших значений кода. Рекомендуется, чтобы отмененные или исключенные элементы кода не присваивались вновь новым элементам закодированного набора.

Приложение ДА
(справочное)

Сведения о соответствии ссылочных международных стандартов национальным стандартам

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ISO/IEC 646:1991	MOD	ГОСТ 27463—87 «Система обработки информации. 7-битные кодированные наборы символов»
ISO 2382-4:1987	—	*
ISO 3166:1993	MOD	ГОСТ 7.67—2003 (ИСО 3166-1:2006) «Система стандартов по информации, библиотечному и издательскому делу. Коды названий стран»
ISO 7064:1983	—	*
ISO 8601:1988	IDT	ГОСТ ИСО 8601—2001 «Система стандартов по информации, библиотечному и издательскому делу. Представление дат и времени. Общие требования»
ISO 9735:1988	MOD	ГОСТ 6.20.1—90 (ИСО 9735—88) «Электронный обмен данными в управлении, торговле и на транспорте (ЭДИФАКТ). Синтаксические правила»
ISO/IEC 11179-3:1994	—	*

* Соответствующий национальный стандарт отсутствует. До его утверждения рекомендуется использовать перевод на русский язык данного международного стандарта.

Примечание — В настоящей таблице использованы следующие условные обозначения степени соответствия стандартов:

- IDT — идентичные стандарты;
- MOD — модифицированные стандарты.

УДК 35.040.00:006.354 ОКС 35.100.01 П85

Ключевые слова: обработка данных, информационный обмен, элементы данных, представление данных, преобразование данных, кодирование

Редактор *Л.В. Коретникова*
Технический редактор *И.Е. Черепкова*
Корректор *С.И. Фирсова*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 10.12.2018. Подписано в печать 25.12.2018. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 3,26. Уч.-изд. л. 2,95.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ИД «Юриспруденция», 115419, Москва, ул. Орджоникидзе, 11
gost.e-ecolog.ru gost.e-ecolog.ru
Создано в единичном исполнении ФГУП «СТАНДАРТИНФОРМ»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru