# Лабораторная работа №4

**Тема:** «Классификация и кодирование»

**Цель:** научитьсяклассифицировать и кодировать информацию, Создать кодификатор множества объектов, используя возможности электронных таблиц MS Excel.

**Теоретический материал**

*Классификацией* называется логическая операция, которая заключается в распределении элементов рассматриваемого множества по подмножествам (классам) на основании общего признака или группы признаков. Каждый объект классификации характеризуется рядом свойств. То свойство объекта, которое позволяет установить его сходство или различие с другими объектами классификации, называется признаком классификации.

Упорядоченное расположение классифицируемых элементов на основе установленных связей и зависимостей между их признаками представляет собой классификационную систему.

*Классификатор* – систематизированный свод наименований и кодов классификационных группировок.

При любой классификации желательно, чтобы соблюдались следующие требования:

* полнота охвата объектов рассматриваемой области;
* однозначность реквизитов;
* возможность включения новых объектов.

В любой стране разработаны и применяются государственные, отраслевые, региональные классификаторы. Так, например, классифицированы: отрасли промышленности, оборудование, профессии, единицы измерения, статьи затрат и т.д.

Существует несколько разновидностей классификаций: иерархическая, фасетная и дескрипторная.

Под иерархией понимается такое соотношение, когда один класс является подклассом другого, целиком объемлющего первый.

Если в классификации каждое множество высшего порядка содержит непересекающиеся между собой подмножества низшего порядка, то эта классификация называется иерархической. Иногда такая классификация называется линейной. Например, деление людей на мужчин и женщин. Исходное множество – «люди» делится на непересекающиеся между собой подмножества: «мужчины» и «женщины». Здесь основанием деления является пол человека.

Иерархическая классификация строится следующим образом: берется исходное понятие (множество) и последовательно уменьшается его объем (последовательно разбивается на подмножества) в соответствии с рисунком 1.1.



Рисунок 1.1 – Пример иерархической классификации

Под объемом понятия понимается множество предметов, каждый из которых обладает всеми признаками, составляющими содержание данного понятия

В иерархической системе классификации из-за жесткой структуры особое внимание следует уделить выбору классификационных признаков.

*Пример*. Создать иерархическую систему классификации для информационного объекта «Верхняя одежда», которая позволит классифицировать информацию об объекте по следующим классификационным признакам: вид, тип и материал изготовления. Сведем основные признаки в таблицу №1.1.

Таблица №1.1 – Значения признаков классификации верхней одежды

|  |  |  |
| --- | --- | --- |
| Признак | Уровень классификации | Значения признака |
| Вид | 1 | * Пальто * Шуба * Куртка * Плащ |
| Тип | 2 | * Длинное * Короткое |
| Материал | 3 | * Драп * Кашемир * Шерсть * Натуральный мех * Искусственный мех * Кожа * Ткань |

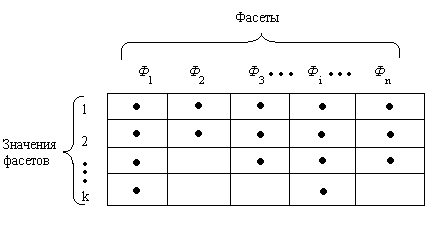
*Фасетная система классификации*, в отличие от иерархической, позволяет выбирать признаки классификации независимо как друг от друга, так и от семантического содержания классифицируемого объекта. Признаки классификации называются *фасетами* (facet – рамка). Каждый фасет содержит совокупность однородных значений данного классификационного признака. Причем значения в фасете могут располагаться в произвольном порядке, хотя предпочтительнее их упорядочение.

Фасетный можно понимать и как «отдельный», то есть каждый признак объекта (фасет) рассматривается индивидуально, а не в связи с остальными признаками. В основе фасетной классификации лежит фасетный анализ. Анализируются характерные признаки объектов классификации и выявляются основные категории свойств предметов. Фасеты в свою очередь могут быть разделены на субфасеты и т.п. до необходимой степени детализации.

*Пример.* Фасет одежды содержит значения мужская, женская, детская. В свою очередь каждую группу можно разбить на зимнюю, летнюю, демисезонную. Каждую подгруппу можно разбить по размеру, цвету и т.д.

После того как все характерные признаки выявлены, их необходимо упорядочить, то есть зафиксировать порядок перечисления этих признаков в ходе описания объекта исходного множества. Схема фиксированной последовательности расположения фасетов в классификации называется *фасетной формулой*.

Схема построения фасетной системы классификации в виде таблицы отображена на рисунке 1.1. Названия столбцов соответствуют выделенным классификационным признакам (фасетам), обозначенным Ф1, Ф2, …, Фi, …, Фn. Например, цвет, размер одежды, вес и т.д. Произведена нумерация строк таблицы. В каждой клетке таблицы хранится конкретное значение фасета.

Рисунок 1.2 – Фасетная система классификации

*Пример*. Разработать фасетную систему классификации верхней одежды с учетом сезонности.

Сгруппируем и представим в виде таблицы все классификационные признаки по фасетам (таблица №1.2).

– фасет вид одежды с четырьмя видами;

– фасет тип с двумя типами исполнения;

– фасет материал с семью наименованиями материала;

– фасет сезонность с двумя обозначениями сезонов.

Структурную формулу любого класса можно представить в виде:

Ks=(вид одежды, тип, материал, сезонность).

Присваивая конкретные значения каждому фасету, получим следующие классы:

К1=(куртка, короткая, кожа, демисезонная);

K2=(шуба, длинная, искусственный мех, зимняя) и т.д.

Таблица №1.2 – Фасетная система классификации

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Вид | Тип | Материал | Сезонность |
| 1 | Пальто | Длинная | Драп | Зимнее |
| 2 | Шуба | Короткая | Кашемир | Демисезонное |
| 3 | Плащ |  | Шерсть |  |
| 4 | Куртка |  | Натуральный мех |  |
| 5 |  |  | Искусственный мех |  |
| 6 |  |  | Кожа |  |
| 7 |  |  | Ткань |  |

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

Суть дескрипторного метода классификации заключается в следующем:

– отбирается совокупность ключевых слов или словосочетаний, описывающих определенную предметную область или совокупность однородных объектов. Причем среди ключевых слов могут находиться синонимы;

– выбранные ключевые слова и словосочетания подвергаются нормализации, т.е. из совокупности синонимов выбирается один или несколько наиболее употребимых;

– создается словарь дескрипторов, т.е. словарь ключевых слов и словосочетаний, отобранных в результате процедуры нормализации.

*Пример*. В качестве объекта классификации рассматривается успеваемость студентов. Ключевыми словами могут быть выбраны: оценка, экзамен, зачет, преподаватель, студент, семестр, название предмета. Здесь нет синонимов, и поэтому указанные ключевые слова можно использовать как словарь дескрипторов. В качестве предметной области выбирается учебная деятельность в высшем учебном заведении. Ключевыми словами могут быть выбраны: студент, обучаемый, учащийся, преподаватель, учитель, педагог, лектор, ассистент, доцент, профессор, коллега, факультет, подразделение университета, аудитория, комната, лекция, практическое занятие, занятие и т.д. Среди указанных ключевых слов встречаются синонимы, например: студент, обучаемый, учащийся; преподаватель, учитель, педагог; факультет, подразделение университета и т.д. После нормализации словарь дескрипторов будет состоять из следующих слов: студент, преподаватель, лектор, ассистент, доцент, профессор, факультет, аудитория, лекция, практическое занятие и т.д.

Под *кодированием* понимается процесс присвоения условных обозначений названиям (признакам) объектов. Цель кодирования – устранение неопределенности и представление информации в более компактной форме для обработки на ЭВМ. Система кодирования – совокупность правил кодового обозначения объектов. Каждый код в своей номенклатуре должен быть единственным для конкретной позиции.

*Порядковый метод*

Порядковый метод кодирования наиболее простой. Суть его заключается в том, что позиции кодируемой номенклатуры обозначаются порядковыми номерами. Предварительной классификации при этом не требуется. Расположение позиций номенклатуры для их кодирования может проводиться в любом порядке: хронологическом, алфавитном и др.

*Серийно-порядковый* метод отличается от порядкового тем, что номенклатура кодируемых объектов предварительно разбивается на подмножества. Эти подмножества состоят, как правило, из элементов, обладающих общим свойством. Элементы каждого подмножества кодируются порядковыми номерами в пределах отведенной для него серии номеров. По своей сути серийно-порядковая система является смешанной: классифицирующей и идентифицирующей. Применяется тогда, когда количество подмножеств невелико.

Пример. Все студенты одного факультета разбиваются на учебные группы (в данной терминологии – серии), для которых используется порядковая нумерация. Внутри каждой группы производится упорядочение фамилий студентов по алфавиту, и каждому студенту присваивается номер.

В каждой серии обычно предусматриваются резервные номера (на случай появления новых позиций), которые могут располагаться в любом месте серии.

*Пример*. Закодируем оборудование ателье по пошиву одежды.

В приведенном на этом примере коде серия номеров 01-19 обозначает мебель (07-19 – резервные коды), серия 20-39 обозначает оборудование (26-39 – резервные коды), 40-59 обозначает инструменты (45-59 – резервные коды), а 60-99 – материалы (65-99 – резервные коды).

*Комбинированные* методы кодирования представляют собой синтез вышеописанных методов кодирования. Они используются для кодирования больших номенклатур объектов, которые можно группировать по нескольким соподчиненным или независимым признакам.

**Порядок выполнения работы**

Прежде чем составлять кодификатор, необходимо классифицировать множество возможных изделий. Для этого нужно выбрать необходимый перечень признаков, по которым будем делить исходное множество на подмножества и определимся со способом классификации.

Признаки классификации представлены в таблице № 1.3.

Таблица №1.3 – Признаки классификации

|  |  |  |
| --- | --- | --- |
| № п/п | Признак классификации | Расшифровка признака |
| 1 | Пол | Половой признак |
| 2 | Тип изделия | Подмножество: рубашка, пиджак, платье и т.д. |
| 3 | Вид изделия | Детализация типа изделия |
| 4 | Модель изделия | Конкретная модель вида изделия |
| 5 | Усложненность модели | Элементы, усложняющие исходную модель |
| 6 | Размер изделия | Размер изделия |

Итак, выбираем необходимые признаки, по которым будем классифицировать объекты исходного множества. В связи с отсутствием жесткой иерархии подчиненности выбираем фасетную классификацию.

Теперь необходимо разработать код, однозначно определяющий каждое конкретное изделие. Поскольку выбранная классификация фасетная, то кодировать будем, используя позиционный параллельный метод кодирования. Количество фасетов определяется количеством признаков, выбранных для деления исходного множества на подмножества.

Таким образом, разрабатываемый код будет иметь шесть фасетов. Выполним их наполнение для определения разрядности кода по каждому фасету.

*Наполнение фасет*

Произведём наполнение фасет возможным перечнем номенклатуры, присваивая каждому фокусу свой код. Кодирование фасет будем производить порядковым или серийно-порядковым методом (таблицы 1.4, 1.5).

Таблица №1.4 – Фасет «Пол»

|  |  |
| --- | --- |
| Наименование позиции номенклатуры | Код |
| Мужской | 1 |
| Женский | 2 |

Таблица №1.5 – Фасет «Тип изделия»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Для значения 1 фасета «Пол» | |  | Для значения 2 фасета «Пол» | |
| Наименование позиции номенклатуры | Код |  | Наименование позиции номенклатуры | Код |
| Шуба | 01 |  | Шуба | 01 |
| Пальто | 02 |  | Пальто | 02 |
| Плащ | 03 |  | Плащ | 03 |
| Куртка | 04 |  | Куртка | 04 |
| Жилет | 05 |  | Жилет | 05 |
| Комбинезон | 06 |  | Жакет | 06 |
| Костюм спортивный | 07 |  | Комбинезон | 07 |
| Рубашка | 08 |  | Костюм спортивный | 08 |
| Брюки | 09 |  | Рубашка | 09 |
| Костюм | 10 |  | Брюки | 10 |
| Пиджак | 11 |  | Костюм | 11 |
|  |  |  | Юбка | 12 |
|  |  |  | Блузка | 13 |
|  |  |  | Пиджак | 14 |
|  |  |  | Платье летнее | 15 |
|  |  |  | Платье вечернее | 16 |
|  |  |  | Платье свадебное | 17 |
|  |  |  | Сарафан | 18 |

*Вид изделия.*Кодирование этого фасета произведем серийно-порядковым кодом. Приведем наполнение фасета на примере кодов 03 «Плащ», 12 «Юбка» и 14 «Пиджак» типа изделия для значения 2 кода пола (таблица № 1.6).

Таблица 1.6 –Фасет «Тип изделия» для кодов 03,12, 14

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 03 «Плащ» | | 12 «Юбка» | | 14 «Пиджак» | |
| Наименование позиции номенклатуры | Код | Наименование позиции номенклатуры | Код | Наименование позиции номенклатуры | Код |
| Плащ удлиненный | 01 | Юбка макси | 01 | Пиджак двубортный | 01 |
| Плащ укороченный | 02 | Юбка мини | 02 | Пиджак однобортный | 02 |
| Плащ средней длины | 03 | Юбка средней длины | 03 | Пиджак приталенный | 03 |

*Модель изделия***.** Заполнение фасета по этому признаку рассмотрим на примере кода 12 «Юбка» типа изделия, разбитого на подмножества применением кода вида изделия (таблица № 1.7).

Таблица № 1.7 – Фасет «Модель изделия» для кода 12

|  |  |
| --- | --- |
| Наименование позиции номенклатуры | Код |
| «годе» | 01 |
| прямая | 02 |
| «солнце» | 03 |
| татьянка | 04 |
| 3 клина | 05 |
| 4 клина | 06 |
| 6 клиньев | 07 |
| 8 клиньев | 08 |
| 12 клиньев | 09 |
| трапеция | 10 |

*Усложненность модели.*Кодирование проведем серийно-порядковым кодом, предусматривая резерв для увеличения номенклатуры (таблица № 1.8).

Таблица № 1.8 – Фасет «Усложненность модели»

|  |  |
| --- | --- |
| Наименование позиции номенклатуры | Код |
| накладной карман | 001 |
| врезной карман | 002 |
|  |  |
| без пояса | 011 |
| шлевки | 012 |
| фигурный пояс | 013 |
| резинка | 014 |
|  |  |
| застежка брючная | 030 |
| застежка во всю длину пуговицы | 031 |
| застежка во всю длину молнии | 032 |
|  |  |
| с разрезами по бокам | 051 |
| со шлицей | 052 |
| с разрезом | 053 |

В приведенном примере серия 001-010 обозначает карманы (резервные номера 003-010), серия 011-029 – вид пояса (резервные номера 015-029), серия 030-050 – застежка (резервные номера 034-050) и т.д. При необходимости номенклатура может быть расширена за счет избыточной емкости кода.

*Размер***.** Код размера изделия совпадает с размером изделия (таблица № 1.9).

Таблица 9 – Фасет «Размер»

|  |  |
| --- | --- |
| Наименование позиции номенклатуры (размер) | Код |
| 38 | 38 |
| 40 | 40 |
| 42 | 42 |
| . . . . . . . . . . . . . . . . . | |

*Установка разрядности признаков*

Установка разрядности признаков производится по порядковым номерам фасетов. Исходя из числа возможных объектов номенклатуры по каждому признаку, сведем разрядность признаков в таблицу № 1.10.

Таблица № 1.10 – Разрядность признаков

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Признак | 1 | 2 | 3 | 4 | 5 | 6 |
| Количество разрядов кода | 1 | 2 | 2 | 2 | 3 | 2 |

Итак, в структуре двенадцатиразрядного кода (1+2+2+2+3+2) содержится шесть фасетов.

***Формирование фасетной формулы***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | XX | XX | XX | XXX | XX |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  | Размер изделия |
|  |  |  |  |  |  | Усложненность изделия |
|  |  |  |  |  |  | Модель изделия |
|  |  |  |  |  |  | Вид изделия |
|  |  |  |  |  |  | Тип изделия |
|  |  |  |  |  |  | Пол |

Примеры кода

Женская юбка средней длины из 3-клиньев со шлицей 48 размера будет иметь код 212030505248.

Женская юбка средней длины из 4 клиньев без пояса 44 размера будет иметь код 212030601144.

Женская юбка мини вида трапеция с накладными карманами с брючной застежкой на поясе со шлевками 44 размера.

Для реализации кодификатора в среде MS Excel необходимо табличный материал списка закодированных изделий со всеми признаками перенести на лист программы MS Excel. Далее, используя встроенные функции MS Excel из раздела Текстовые: =ПСТР(), =ЗНАЧЕН() и встроенную функцию из раздела Ссылки и массивы: ВПР() необходимо организовать расшифровку кода для составленного в Лабораторной работе №1 кода.

Поскольку фасеты по всем выбранным признакам были заполнены только для позиции «Юбка», то и список будет составлен только для этой позиции, так как для остальных позиций разработанного кодификатора фасеты заполнены не по всем признакам.

Встроенная функция =ПСТР() возвращает заданное число знаков из текстовой строки, начиная с указанной позиции. Этой функцией воспользуемся для извлечения из общего кода отдельных цифр кода для их расшифровки.

Встроенная функция =ЗНАЧЕН() используется для преобразования строки текста, отображающую число, в число.

Комбинация встроенных функций и возможность вложенности одной функции в другую позволяют извлечь отдельные цифры кода из общего кода в соответствии с рисунком 2.1.



Рисунок 2.1 –Извлечение отдельных цифр кода

Внешняя встроенная функция ЗНАЧЕН() позволяет преобразовать отдельные цифры кода в числовой формат, поскольку в общем коде программа распознавала цифры кода текстом. Таким образом, организовав работу комбинации встроенных функций =ЗНАЧЕН(ПСТР()) получим в отдельных ячейках отдельные цифры общего кода. Это необходимо для расшифровки по таблицам сведений, закодированных в общем коде.

Встроенная функция =ВПР() позволяет выполнить поиск в первом столбце [диапазона](javascript:AppendPopup(this,'xldefRange_2_2')), а затем вернуть значение из любой ячейки в той же строке диапазона. Важным моментом является расположение цифр кода и текстовых расшифровок: цифры кода должны находиться в левом столбце, а текст расшифровок в правом столбце в соответствии с рисунком 2.2.

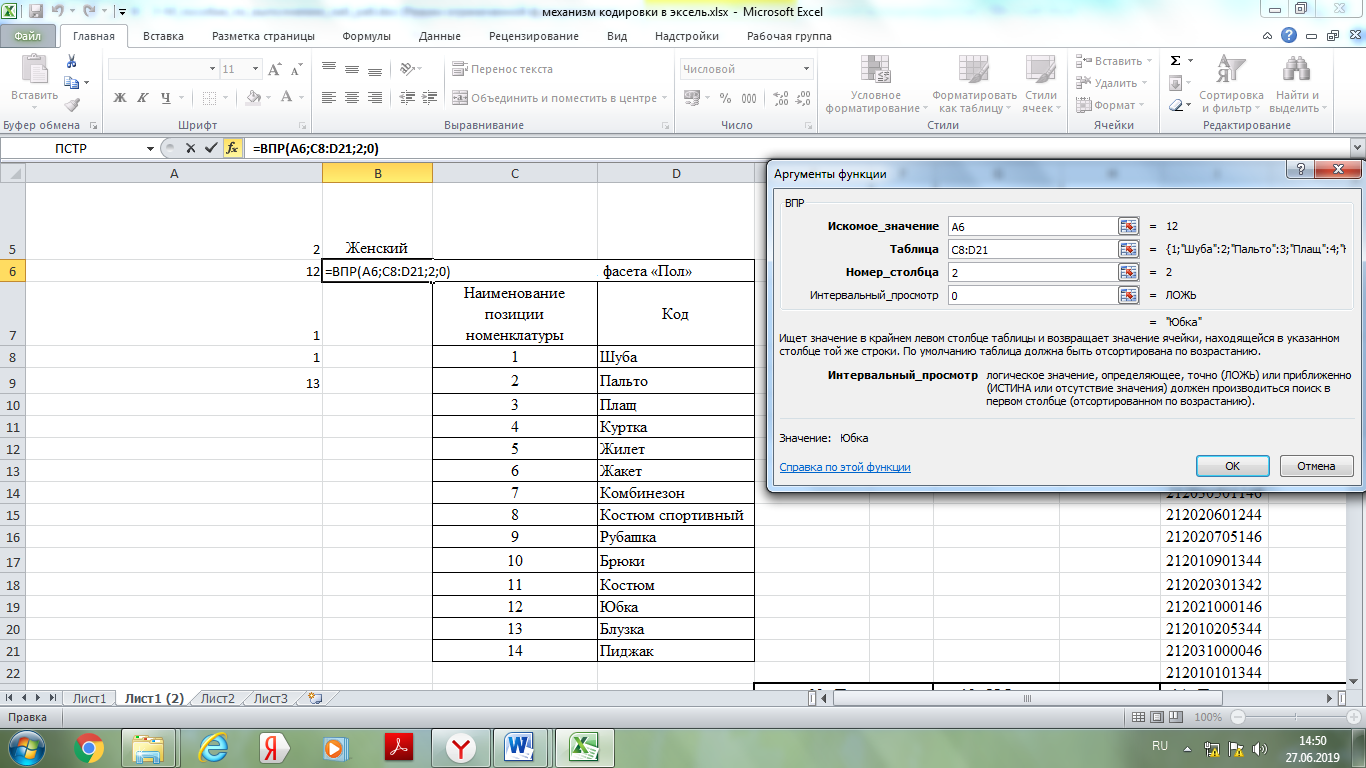


Рисунок 2.2 – Использование встроенной функции ВПР()

Результат представлен на рисунке 2.3. Таким образом, необходимо организовать работу встроенной функции для каждой цифры кода и расшифровать весь код по всем таблицам.

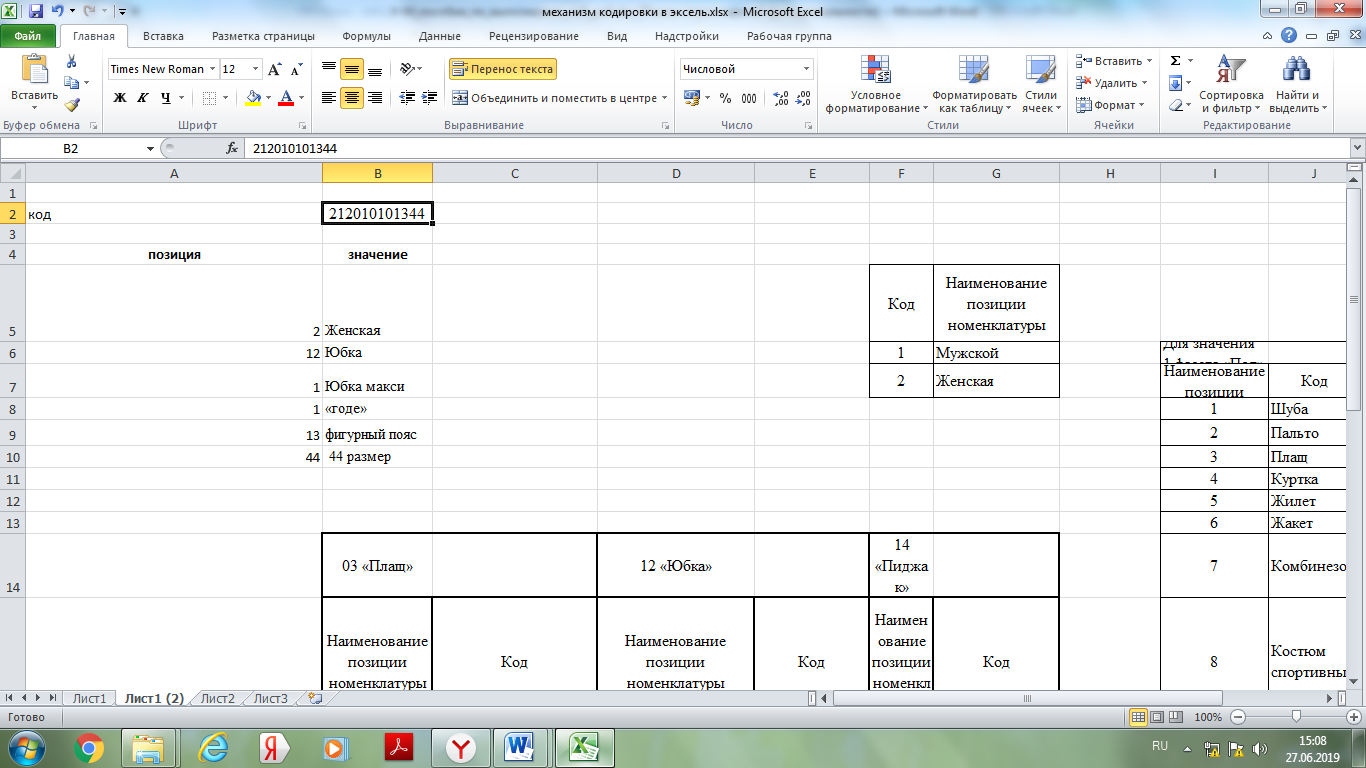


Рисунок 2.3 – Расшифровка кода по исходным таблицам

**Задание для лабораторной работы.**

**Задание 1.** Создать классификатор и кодификатор для индивидуальной предметной области.

**Варианты:**

1) Классифицировать и закодировать моделей процессов современных персональных компьютеров.

2) Классифицировать и закодировать множество возможных изделий класса верхней одежды с целью получения возможного перечня моделей каждого вида.

3) Классифицировать и закодировать множество видов и разновидностей микросхем оперативной памяти.

4) Составить классификатор и закодировать реестр регистрационных карт медицинского учреждения.

5) Классифицировать и закодировать ассортимент магазина по продаже комплектующих персонального компьютера.

6) Классифицировать и закодировать множество видов и разновидностей компьютерных мышей.

7) Классифицировать и закодировать множество видов и разновидностей принтеров.

8) Классифицировать и закодировать множество видов и разновидностей сканеров.

9) Классифицировать и закодировать множество видов и разновидностей применяемых современных гаджетов.

10) Классифицировать и закодировать любой из видов мониторов.

**Задание 2.** Для индивидуального варианта предметной области организовать работу по расшифровке трех кодовых позиций с помощью встроенных функций в среде MS Excel.

**Содержание отчета:**

Тема работы.

Цель работы.

Задание к работе.

Распечатка материала выполнения лабораторной работы.