UDC 004.4

THE USE OF CODING METHODS AND SYSTEMS IN THE DEVELOPMENT OF ENCODING CHARACTERS FOR TEXT INFORMATION IN DATABASES

ИСПОЛЬЗОВАНИЕ МЕТОДОВ И СИСТЕМ КОДИРОВАНИЯ ПРИ РАЗРАБОТКЕ КОДИРУЮЩИХ СИМВОЛОВ ДЛЯ ТЕКСТОВОЙ ИНФОРМАЦИИ В БАЗАХ ДАННЫХ

Stasenia T., [1]Mandrik O.
Vitebsk State Technological University, Republic of Belarus
E-mail: [1]mandrik_miit@rambler.ru
Стасеня Т.П., Мандрик О.Г.
Витебский государственный технологический университет, Республика Беларусь

ABSTRACT

DATABASES, DATA TYPES, SOFTWARE

Today, to increase productivity and significantly reduce costs, automatic identification technologies are used, the main ones being the following: bar coding; radio frequency systems; optical character recognition; machine vision (computational methods of image processing); speech data entry and a number of others. Encoding characters provide quick access to the characteristics of the object and simplify the processing of text information.

АННОТАЦИЯ

БАЗЫ ДАННЫХ, ТИПЫ ДАННЫХ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Сегодня для повышения производительности труда и существенного снижения затрат используются технологии автоматической идентификации, основные из них: штриховое кодирование; радиочастотные системы; оптическое распознавание знаков; машинное зрение (вычислительные методы обработки изображения); речевой ввод данных и ряд других. Кодирующие символы обеспечивают быстрый доступ к характеристикам объекта и упрощают обработку текстовой информации.

For effective production management, product quality improvement and organization of accounting there must be timely and accurate control of objects.

The primary collection of information requires the development of a system of identification of various types of objects (goods, documents, tools, etc.). To solve this problem, it is necessary to determine the systems and methods of identification of objects or three-dimensional text characteristics.

Many of the characteristics of objects are text information. Text information is difficult to convert to binary code because of the variety of characters and length. Each character in the text is encoded in one byte. A lot of automation resources are spent on translation and restoration of text information. Working with this type is difficult: distortion when entering; inaccuracies when adjusting; character-by-character work when searching and comparing.

For the rational operation of software with long texts, coding is used. Encoding characters replace many text characteristics and make it easier to process text data. In everyday life, we often encounter coded information, for example, the number of the group of students replaces the following information: the code and the full name of the specialty, faculty, course; student ID number (the grade book) uniquely identifies the student among all students of the Republic of Belarus. It often happens that the encoded information is more convenient to familiarize with the amount of data and is better perceived visually.

When developing an information-logical database model, encoding characters are actively used for field names, individual objects and/or object characteristics. Regardless of the direction of the developed SOFTWARE (software), the basic requirements for information are: adequacy, reliability, completeness, objectivity and accuracy.

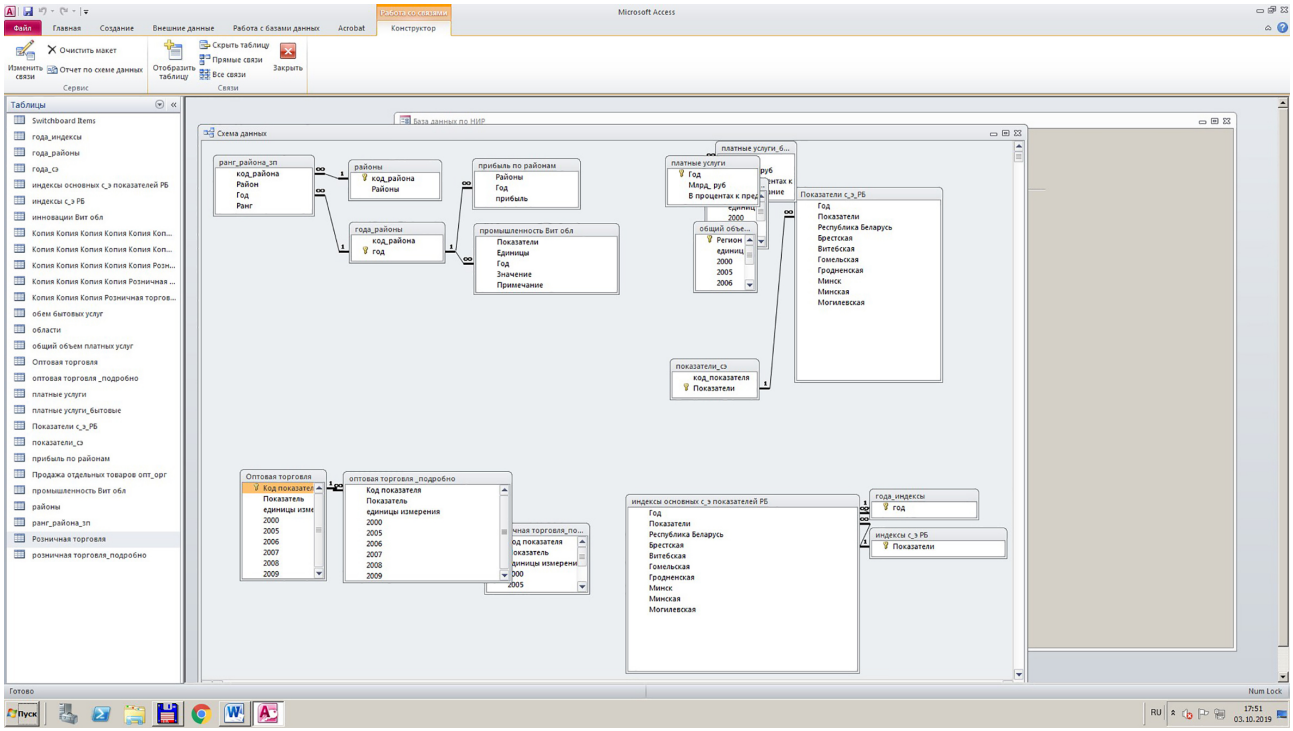

Figure 1 – Fragment of the data scheme using the area code field (in the development of the database of economic indicators of Vitebsk region)

Development of code and code directories begins with the initial collection of information in the design of the information system. Often there is a need to assign each object a number

(code), followed by the application of this number on the object for reading by digital devices. To streamline the coding symbols used by the classifiers. When encoding text information as a basis, you can take existing classifiers, which are known in many fields of activity.

When designing the information system software, the following actions are performed:

– definition of the list of necessary indicators;

– calculation of space-time characteristics of indicators;

– identification of existing information links;

– development of database structure and composition;

– study of existing classifiers: industry, government, etc.;

– development of new codes and classifiers or adaptation of existing ones;

– structural organization and visual design of forms for different purposes.

To logically describe data structures, identifiers are used, i.e. symbols that characterize individual units of information and show their place, level and relationship in the overall system. The purpose of identification is a simplified description of data and processing algorithms. There are systemic, structural and procedural ways of identification. At system identification the full list of information units is made: details, indicators, forms of documents to which symbols on all positions at the level of system are appropriated; at structural level each information unit on the basis of its coordinates in structure of documents or carriers; at procedural – within separate block diagrams, modules and programs.

Special graphical symbols of operations are used to describe information transformation procedures and algorithms.

Encoding of information has always been paid attention to when working with large arrays of objects. Two methods of classification are developed: hierarchical and faceted. There are four encoding methods: sequential, batch-sequential, sequential, parallel. Each method when used has its positive and negative sides.

The ordinal method is characterized by its simplicity, but it does not allow you to skip the encoding numbers and insert the encoding characters into the formed series.

The serial-order method allows extensions of certain groups to be performed, which is sometimes necessary when the number of objects increases, but is not able to preserve the structure when the number of groups increases.

The sequential method is characterized by hierarchical organization of data with strictly defined features (characteristics) of objects at each level. This method is effective when using large groups of objects of different types. In programming, this structure is used in class diagrams.

When working with large items, it is more rational to use the parallel method. It is necessary to develop a coding structure, which is a disadvantage.

The choice of coding method is influenced by the purpose and conditions of use of the information system. The structure of the code should be simple and logically clear, because it

may be necessary to automate the coding process. The coding system should provide for an increase in the list of objects and the expansion of the set of characteristics characteristic of the operating conditions of the software. Typically, an electronic information system provides a mechanism for the development of validation codes.

The method and structure of the formation of code symbols of the object or feature should be reflected in electronic directories, develop explanations and reference information for users. The local coding system should not create conflicts and mislead users, if there are already established encoding character combinations in this area.

The historical system can be taken as a basis and refined, even if it contradicts the principle of efficiency. In this situation, it is very rational to approach the assessment of the situation, because users of the system should not work for the sake of software, and it should provide them with convenient and comfortable working conditions with information.

Local classifiers with stable information are electronic directories. The stability coefficient is determined by the ratio of the amount of changed information to the total amount of data for a certain period. According to the coefficient of stability, if it is greater than 0.5, the normative reference information for its repeated use is presented in electronic form.

If you need a stable storage of information for a long period of requirements for the encoding mechanism increases significantly. Thoughtful and high-quality solution for encoding text or volume information simplifies the design and commissioning of the next version of the information system.

It is not difficult to imagine how effective the use of automated input of initial information in enterprises, wholesale bases, supermarkets. Coding of objects using computer technology makes it possible to organize effective work on the processing of interconnected arrays of information.

In the study we can conclude that the design of automated information processing it is necessary to encode text data types and to develop a mechanism for compiling a synthesis of code names for selected volumes of information. Identification of objects by code allows: rational use of SOFTWARE and hardware capabilities; organize effective control and accounting; avoid losses or inaccuracies in the output, analysis or editing of text characteristics. In the economic sphere, a quick search, counting, selection of goods or the output of all information about the product can improve the quality of customer service; optimally perform warehouse work; efficiently use vehicles; perform inventory control and reduce the time of work with documents.