УЧЕБНИК

В.П. Омельченко, А.А. Демидова

МЕДИЦИНСКАЯ ИНФОРМАТИКА

2-е издание,
переработанное

Министерство образования и науки РФ

Рекомендовано ГБОУ ДПО «Российская медицинская академия последипломного образования» Минздрава России к использованию в качестве учебника в образовательных учреждениях, реализующих программы ВПО по дисциплине «Медицинская информатика» по базовым медицинским специальностям

TEXTBOOK

V.P. Omelchenko, A.A. Demidova

MEDICAL INFORMATICS

Moscow
«GEOTAR-Media»
PUBLISHING GROUP
2020

CONTENTS

LIST OF ABBREVIATIONS

AWP	— automated workplace
AWS	— automated working station
ADC	— analogue/digital converter
DB	— database
KB	— knowledge base
BFB	— biological feedback
WHO	— World Health Organisation
VHI	— voluntary health insurance
UPHIS	— Unified Public Health Information System
CHD	— coronary heart disease
AI	— artificial intelligence
ICT	— information computer technologies
IS	— information system
IT	— information technologies
IEMR	— integrated electronic medical record
FDCS	— functional diagnostic computer system
CT	— computer tomography
LAN	— local area network
TDP	— therapeutic & diagnostic process
LIS	— laboratory information system
HCP	— health care provider
RG	— rehabilitation gymnastics
MIS	— medical information system
MO	— medical organisation
MP	— microprocessor
MCDS	— medical device and computer system
HDD	— hard disk drive
OD	— optical drive
RAM	— random access memory
CHI	— compulsory health insurance
OS	— operating system
ROM	— read-only memory
PC	— personal computer
SW	— software
VLSIC	— very large integrated circuit
ISS	— information security system
AIS	— artificial intelligence system
MDSS	— medical decision support system

DBMS	— database management system
DC	— data centre
EDPM	— electronic data processing machine
ECG	— electrocardiography/electrocardiogram
EMR	— electronic medical record
ES	— expert system
EEG	— electroencephalogram
NMR	— nuclear magnetic resonance
API	— Application Programming Interface
BIOS	— Basic Input/Output System
CMOS	— complementary-symmetry/metal-oxide semiconductor
DNS	— Domain Name System
FAT	— File Allocation Table
FTP	— File Transfer Protocol
HL7	— Health Level 7 Standard
HTML	— HyperText Markup Language
HTTP	— HyperText Transfer Protocol
IEEE	— Institute of Electrical and Electronics Engineers
IT	— information technology
LCD	— Liquid Crystal Display
POP3	— Post Office Protocol
RAM	— Random Access Memory
SMTP	— Simple Mail Transfer Protocol
TCP/IP	— Transmission Control Protocol/Internet Protocol
UDP	— User Datagram Protocol
UPnP	— Universal Plug and Play
URL	— Uniform Resource Locator
WWW	— World Wide Web
XML	— eXtensible Markup Language

PREFACE

The 3rd-Generation Federal State Educational Standard (2012) introduced Medical Informatics as a part of the curriculum for first-year students of medical universities. Before that, Computer Science was taught for first-year students, while junior and senior students had courses in Medical Informatics. Such changes in the curriculum make it difficult to teach Medical Informatics for first-year students, as they do not have sufficient training in the field of clinical disciplines and organisation of health care, which are the subject of Medical Informatics.

The list of professional competences of a graduate of a medical school includes the «ability and readiness to work with the medical and technical equipment that is used to treat patients, firm knowledge on how to use computers, how to obtain information from various sources, how to work with information in global computer networks, how to apply the capacities of modern information technologies to solve daily professional tasks» (PK-9).

This textbook is dedicated to the study of these professional competencies. It is based on the approximate curriculum of Medical Informatics prepared under the direction of T. Zarubina, Head of Department of Cybernetics and Informatics of Pirogov Russian National Research Medical University.

Taking into account the different backgrounds of high school graduates in the field of Informatics, the proposed textbook covers both General Informatics, which is partially studied in secondary school, and medical information systems used in the diagnostic and treatment process, and medical institutions management. In addition, the possibilities of Internet resources and the use of telemedicine technology to improve the level of health care is described. The book discusses the idea of creating a Unified Medical Information Space as set out in the Concept of Creating a Unified Public Health Information System» (Order of the Ministry of Health of Russia No. 364 of April 28, 2011).

Chapter 1 defines information and Informatics as a science. It describes the subject and the goals of Informatics. It also presents a classification of medical knowledge and medical documents used by health care providers. The chapter introduces the concept of information technologies and their use in health care.

Chapter 2 is dedicated to hardware and software of computing systems. It includes a classification of EDPMs as well as the structure of a personal computer and the characteristics of the following principal components: the processor, internal and external drives, input and output devices. The functions of system and application programs, as well as programming environments are described in it. An overview of the Windows operating systems is provided. Methods and tools for protection against unauthorised access are given consideration.

Chapter 3 studies the use of MS Office programs by medical staff to solve their professional tasks. Special attention is paid to such applications as MS Word, MS Excel, MS Access and creation of presentations using MS PowerPoint.

Chapter 4 discusses the issues of modelling in health care. It presents the definition of a model and a classification of models (including those used in health care). A detailed description and stages of the construction of mathematical models are described in detail. Examples of mathematical models are given to demonstrate the possibilities of modelling in health care. The features of structural and simulation modelling in health care are studied.

Chapter 5 is dedicated to the development and operation of medical information systems (MISs). Types of MISs are presented, their principles and stages of creation are listed, and the organisation of automated medical staff workplaces is defined. The «Karelian Medical Information System» is used as an example to discuss the functional capabilities of its subsystems.

Chapter 6 discusses the model of the therapeutic and diagnostic process. Automation of the therapeutic and diagnostic process with information and smart support for medical staff significantly increases the efficiency and promptness of health care providers' work. The possibilities of using medical expert systems are considered in this chapter.

Chapter 7 introduces the students to medical devices and computer systems (MDCSs) and their use for the examination, treatment, and rehabilitation of patients. A classification of MDCS is included with their design, main features and examples of commercially available systems.

Chapter 8 discusses laboratory information systems intended for automation of laboratory staff work, the use of automated laboratory analysers, an efficient organisation of laboratory work, reducing manual operations.

Chapter 9 is dedicated to the use of information systems for health management at the municipal, territorial and federal levels. It considers the implementation of the Concept of Creating a Unified Public health care Information System as well as the issues of information security and protection of information in MISs at different levels. A regional MIS is used as an example.

Chapter 10 is dedicated to network technologies in information processing. It reviews topologies, hardware, and software of local area networks and their connection to the Internet. The global Internet network, medical information resources, and search engines are considered in detail. A definition of telemedicine and its essential tools is presented. Examples of creation and use of telemedicine centres for providing health care to the population are presented.

The authors express their gratitude to the staff of the Department of Medicine and Biological Physics of Rostov State Medical University of the Ministry of Health of Russia for the invaluable aid and feedback during the development on this textbook. Special thanks to Associate Professor N.A. Alexeeva for her assistance in preparing the section on Information and Intellectual Support of Therapeutic and Diagnostic Process.

Chapter 1

INFORMATION AND INFORMATION PROCESSES. METHODS AND TOOLS OF ITS DEVELOPMENT IN MEDICINE AND HEALTH CARE

1.1. INFORMATION AND ITS PROPERTIES

Definition

In Latin, information means an explanation, statement of something or description of something. **Information** is data about our environment that reduces the incompleteness of knowledge about the objects and events in it. **Information** is a collection of data that determines the extent of our understanding of certain events, phenomena or factors.

The concept of information, along with matter and energy, is one of the fundamental concepts of the universe, so it is tough to define it accurately.

Concerning computer data processing, information is understood as a particular sequence of symbolic values (letters, numbers, coded graphic images and sounds, etc.) that possesses a precise meaning and is presented in a computer-readable format. Each new symbol in such a sequence increases the information volume of the message.

Information acts as an ability of objects and phenomena (processes) to generate a variety of conditions that are transmitted using the reflection from one object to another. Data covers all aspects and all branches of social life, which is an intricate part of every person's life, affecting their way of thinking and behaviour. It provides communication between people, social groups, classes, nations, and countries, helps people to develop a scientific world outlook, to understand the diverse phenomena and processes of social life, to improve their level of culture and education, learn and abide by the laws and moral principles. The role of information in management is immense and irreplaceable. In fact, if there is no information, it is impossible to speak about any management or purposeful activity of interrelated objects and systems.

The definition of information includes such concepts as *signal*, *data*, *information*, and *knowledge*.

A *signal* is a time-dependent physical process that reflects the specific characteristics of an object. The propagation of a signal is completed by interaction with physical bodies, which is called signal registration. This is when data form. *Data* is the properties of objects registered on a medium that can be measured or compared with certain standards.

Information is data that has been perceived (understood) by a subject (person) and can be used in their (professional) activity. Therefore, information can be defined as the data that is used.

Knowledge is information about an object that is systematically confirmed through experiment or logic.

Thus, the overall scheme of information processes can be represented, as shown in fig. 1.1.

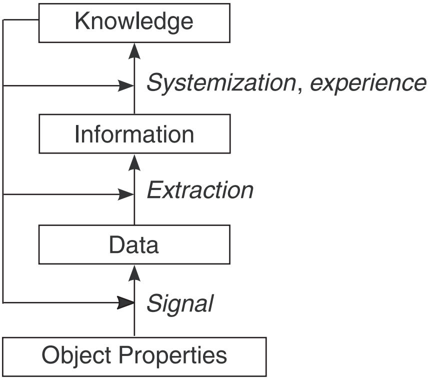

Fig. 1.1. General Model of Information Processes

For example, if electrocardiographic (ECG) method is used to study the condition of the cardiovascular system, the heart is the study object, the bioelectric activity of the heart is the signal, the electrocardiogram is the registered signal, i.e. the data. The ECG record gives the cardiologist information about the state of the cardiovascular system. Processed and analysed ECG records and their correlation with the state of the cardiovascular system is knowledge about the heart, which can be transferred to young staff for practical use.

Let’s list the properties of information.

- *Objectivity* and *subjectivity* reflect the adequacy of information extracting methods. Information objectivity means that it always results from the data on the properties of particular *objects*. Subjectivity means that some person (a subject) can extract information from specific data, while someone else will fail to do so. For example, objective information on a patient’s rhythmic activity disorder would be the registered unequal intervals between heartbeats. And subjective information would be the «fluttering» and «freezing» in the chest that the patient feels now and then.
- *Accuracy*: the degree to which the information corresponds to the real condition of the source of information. For example, a medical certificate without data on the applicant’s previous diseases would present inaccurate information.
- *Reliability* is a probabilistic characteristic that describes the correspondence of information to reality. It is secondary to accuracy.

- *Sufficiency* or *completeness* is the information that is necessary to solve a specific problem. For example, the detection of a characteristic white rash on the inner mucous surface of the cheek (Koplik's spots) would be sufficient to diagnose a child with measles.
- *Availability* or *simplicity* is the possibility to perform procedures to obtain and transform information. In informatics, the availability of data is the avoidance of temporary or permanent concealment of information from users who have obtained access rights. For example, the health information contained in an outpatient medical history is available to the patient. The patient can take their medical history from the registry, get acquainted with the information presented there, submit it to a doctor so that the latter can add information to it. However, if the same patient is admitted to a hospital, they won't have access to medical history. After the patient is discharged, they get access to the discharge report.
- *Relevance*: a value that characterises the period between the moment of occurrence of an event and the presentation of information about it. For example, information about how many times the patient coughs per day, their cough characteristics (productive/non-productive, paroxysmal, painful, etc.), the amount of sputum is relevant for the doctor to diagnose the patient while the latter is sick. But a while after the patient was cured, the information about their cough becomes irrelevant.
- *Value*: the degree of usefulness of information for a particular user. For example, information about the nature of the patient's diet is valuable for a nutritionist to draw up recommendations but is of no value at all for a manager who is selling a computer to the same person.

Information processes include any actions performed with the information: sorting, storage, transfer, and processing.

There are the following levels of *information processes*:

- level 1 — information technologies, which include technical means of information support, software and software systems, information factor, intellectual efforts, and human labour;
- level 2 — information systems: sets of information technologies aimed at such processes as collection, processing, storage, retrieval, transmission, and display of information of the subject area;
- level 3 — information resources: sets of corresponding information systems that are studied additionally, also at social and economic standards of description and application.

1.2. DATA ENCRYPTION

Definition

Data encryption means converting information from one form of representation to another. **Decoding** is the restoration of encrypted information.

In EDPMs, information can be presented in two formats: *analogue* and *digital*.

Analogue means a continuous signal that changes proportionally to the change in information, i.e. the data is encoded by a time-dependent voltage or current. This is the way information was presented in *analogue computing systems* (*ANACOMS*). However, such devices were not developed further, mainly due to the low accuracy of their calculations.

The *digital presentation format* is used in *digital computing systems* (*DCC*). These devices encode the information with numbers. Numbers can be used to encrypt different kinds of information: numbers, letters, sounds, and images. Digital computing systems use the *binary numeric system*, which operates using only two numbers: 0 and 1. There are also other numeric systems: octonary, decimal, hexadecimal, and many others. However, the binary system is characterised by a high degree of reliability of information presentation. It is much easier to recognise two conditions (0 or 1) than, say, ten conditions. In living systems, binary coding of information in the form of rest potential and action potential, biological 0 and 1, is also used to transmit data. In the binary number system, it is possible to perform all mathematical operations, just like in the universal decimal number system.

In digital computing systems, two voltage levels are used to encrypt binary symbols. Typically, 1 means a higher voltage (about 5 V), and 0, a lower one (less than 0.8 V).

There are special devices for converting the analogue to digital, and vice versa. Such devices are called respectively *analogue-to-digital converters* (*ADCs*) and *digital-to-analogue converters* (*DACs*). The process of converting continuous signals into the digital format consists of three stages: discretisation, quantisation, and coding.

The discretisation is the process of splitting the signal into separate components taken at equal intervals, the values of which depend on the discretisation frequency (fig. 1.2, *a*).

The quantisation is the measurement of a discrete value of the signal at the moments t_1, t_2, t_3, and presenting them with precise accuracy. The accuracy is determined by the quantisation levels, i.e. the number of levels of splitting the value of signal *y*.

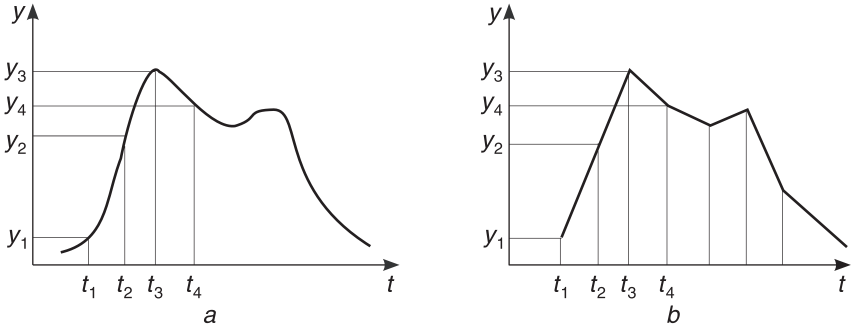

Fig. 1.2. Stages of Discretization (*a*) and reverse conversion of information from digital to analogue (*b*)

Coding is translating the value of the quantisation level into the binary numeric system.

The digital information obtained is called *discrete* information.

DACs perform the reverse process: transform digital signals into *analogue* (fig. 1.2, *b*).

1.2.1. Number coding

Thus, digital computing systems present the information in binary code (i.e. a sequence of 0s and 1s). Each digit is referred to as a bit (*binary digit*). An 8-bit sequence is called a *byte*. A byte can represent a decimal number from 0 to 255, as $2^8 = 256$. Increasing to 16 bit allows encrypting integer numbers between 0 and 65,535 ($2^{16} = 65{,}536$).

In digital computing systems, numbers can be presented in two formats: *fixed-point numbers* and *floating-point numbers* (standard form). In fixed-point numbers, the integer part of the number is separated from the fractional part by a point, for example, 25.386; –0.0025. This is the format that is used for input and output of numeric information.

The floating-point format allows presenting a number more compactly, avoiding using zeroes before and after the point, thus expanding the range of the names that can be used. In its standard form, a number can be presented like:

$$N = \pm M \times 10^{\pm k},$$

with M being the mantissa of the number; k being the power of the number. In this case, the numbers above will look as follows: $+0.25386 \times 10^2$; -0.25×10^{-2}.

1.2.2. Text encryption

Any letter or symbol is represented in the computer as a binary code. The most common one is ASCII (American Standard Code for Information Interchange), which is used for internal representation of character information in the MS-DOS operating system, in the Windows Notepad as well as for coding text files on the Internet. The structure of the code is presented in table 1.1 (the columns and rows are highlighted in bold). The table of codes contains 16 columns and 16 rows; each row and column are numbered in hexadecimal numerals from 0 to F. Hexadecimal presentation of ASCII Code is composed of the number of the row and the number of the column the symbol is located in. 256 characters can be encoded like this.

Table 1.1. ASCII Codes Table

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	...	...		0	@	P	‘	p	A	P	a	...	...	...	p	F
1	...	...	!	1	A	Q	a	q	B	C	b	...	...	...	c	f
2	...	...	«	2	B	R	b	r	C	T	c	...	...	...	t	Є
3	...	...	#	3	C	S	c	s	D	У	d	...	...	...	y	є
4	...	...	$	4	D	T	d	t	E	Ф	e	...	...	...	ф	Ї
5	...	...	%	5	E	U	e	u	F	X	f	...	...	...	x	ї
6	...	...	&	6	F	V	f	v	G	Ц	g	...	...	...	ц	Ў
7	...	...	‘	7	G	W	g	w	H	Ч	h	...	...	...	ч	ў
8	...	...	(	8	H	X	h	x	AND	Ш	and	...	...	...	ш	°
9	...	...	)	9	I	Y	i	y	Й	Щ	й	...	...	...	щ	·
A	...	...	*	:	J	Z	j	z	J	Ъ	j	...	...	...	ъ	·
B	...	...	+	;	K	[	k	{	Л	Ы	л	...	...	...	ы	√
C	...	...	,	<	L	\	1	|	M	Ь	м	...	...	...	ь	№
D	...	...	-	=	M	]	m	}	H	Э	н	...	...	...	э	¤
E	...	...	.	>	N	^	n	~	O	Ю	o	...	...	...	ю	■
F	...	...	/	?	O	_	o	¤	П	Я	п	...	...	...	я	

This table is divided into two parts: columns with numbers from 0 to 7 make up the code standard (the permanent part); columns with numbers from 8 to F are an extension of the code and are used, in particular, to encode the characters of national alphabets. Columns 0 and 1 contain control characters, which are used, in particular, to manage the printer. Columns 2 to 7 contain punctuation signs, arithmetic operations, some service characters, as well as uppercase and lowercase letters of the Latin alphabet. The extension of the code includes line-drawing characters, international characters and other symbols.

In the table above, the Russian alphabet is selected as the national alphabet. If a cell is empty, it means it is not being used, and cells with an ellipsis contain characters that are not displayed intentionally.

Example. Use the ASCII code table to encode the group message using the hexadecimal code view.

Result: A3 E0 E3 AF AF A0 (for simplicity, character codes are separated by spaces), and in binary-decimal code the message will look like:

1010 0011; 1110 0000; 1110 0011; 1010 1111; 1010 1111; 1010 0000.

Currently, there is a new standard based on Unicode, a 16-bit universal code, which allows encoding 65,536 different symbols.

1.2.3. Graphic information encoding

The image on the monitor screen is formed by the glow of points, which are referred to as *pixels* («*Picture Elements*»). The entire set of points making up an image is called a *raster*. The number of pixels on the screen defines the *resolution* of the monitor and can range from 640×480 to a record resolution of 3840×2400. The image quality depends on the size of the pixels and the distance between them. The distance between two adjacent points on the screen is referred to as the *grain*: the smaller it is, the higher the image quality. In high-quality monitors, grain size does not exceed 0.1245 mm (200 dpi). When forming a black-and-white image (for example, for ultrasound images) or in black-and-white video surveillance cameras, each point (pixel) can be of one of 256 shades of grey (from white to black), i.e. 1 byte of video memory is enough to encode the brightness of each point in this case.

Previously, monitors with cathode-ray tube were based on the *principle of decomposition*, which allows obtaining any colour by mixing three colours: red, green and blue. To get a colour pixel, three coloured beams were aimed at one point. This coding system is called RGB (an acronym of the colours used). If eight binary digits are used to represent the brightness of each primary colour, i.e. 3 bytes per dot, it is possible to display 16.5 millions of different colour shades, which is close to the sensitivity of the human eye. This mode of representation of colour graphics is called *True Colour*. Given that True Colour mode requires large amounts of memory, there are other approaches to use. Although the quality of their colour transmission is inferior, they require less memory. For instance, in the *High Colour* mode, 2 bytes are used to transfer the colour of one pixel, which allows displaying more than

65,000 colour shades. There also exists an *index mode*, with the code of each pixel storing the colour index in a special table of colour shades, instead of the colour itself. This mode uses only 1 byte of memory.

The optical effect of liquid crystal elements, which play the role of pixels in LCD (Liquid Crystal Display) screens, is based on the change in the optical polarization of reflected or transmitted light under the action of an electric field. The panel consists of a matrix of cells, each of which is located at the intersection of vertical and horizontal coordinate conductors. A colour LCD sensor has a light filter with three cells per pixel of the image — one for displaying the red, the green and the blue dots. The light wave passes through a liquid crystal cell, with each colour having its cell.

1.2.4. Audio information encoding

Sound is a continuous oscillation and belongs to analogue signals. ADCs are used to input analogue signals into EDCMs (see above). For better signal recording, the sampling frequency must exceed the highest frequency of the signal by 2 times. Given that the highest frequency perceived by the human ear lies in the range between 16 and 20 kHz, the sampling rate of the order of 44 kHz is chosen. The accuracy of measuring the amplitude of the converted signal depends on the bit conversion or quantisation levels of the signal: the more bits, the more accurate the signal digitisation. In practice bit widths of 8, 16 and 24 bits are used. The abovementioned principles of audio encoding are used in the WAV (WAVeform) sound format.

Currently, the MP3 (MPEG Layer 3) format is more popular thanks to its more compact size. MP3 is a streaming format. This means that the data is transmitted by a stream of independent individual blocks of data referred to as frames. To do this, the original signal is divided into sections of equal duration (frames), which are encrypted separately. When decoding, the signal is formed from a sequence of decoded frames. Each frame consists of two granules. A granule consists of two parts that are needed for audio restoration: scale coefficients for each band and a long sequence of Huffman bits. (Huffman algorithm: mathematical data compression algorithm.) After the completion of the two granules, the encoder combines them into a single frame for transmission.

1.2.5. Video encoding

Video information represents the flow of a sequence of images. It is necessary to digitise and store a large amount of information, which is associated with the encoding of the state of each pixel of the screen and the simultaneous recording of the sound. Therefore, high-speed data exchange devices and high-capacity memory storage devices are used. To reduce the amount of information, a unique encoding method characterised by *compression coefficient* is used. The higher the compression ratio, the less memory the information can require, but the lower the image quality. There are several image compression technologies available. The standards here are the developments proposed by MPEG (Monitor Picture Expert Group). In 1999, the MPEG standard was developed, allowing to write a full-length colour film on a regular CD.

In 1999, the MPEG standard was introduced. It is an international compression standard designed for moving objects. The MPEG data compression algorithms reduce the file size so that the data can be transferred more quickly, and then convert them to their original state. Repetitions occurring in adjacent frames are removed, thereby reducing the file size. The compression level can be as high as 50:1. Currently, MPEG includes three compression standard:

- MPEG-1;
- MPEG-2;
- MPEG-4.

Also, MPEG-7 and MPEG-21 standards are being developed.

MPEG-4 (MP4) is a standard for compressing moving images used on the Internet, in radio broadcasting and on information carriers. Compared to MPEG-2, MPEG-4 provides improved quality and smaller file size. In this case, the user is provided with a convenient possibility to save movies on a regular CD, and the quality, in this case, is usually higher than on VCD (Video CD — standard for storing video with sound on CDs).

Presently the Flash Video (FLV) format has gained popularity. It is a video format that is used to transmit data over the Internet. It is the format used on YouTube, Google Video, RuTube, etc. The popularity of this format is due mainly to being supported by Adobe Flash Player. An FLV file is a bitstream that is a variant of the H. 263 video standard commonly referred to as Sorenson Spark. Flash Player 8 and newer editions support On2 TrueMotion VP6 video streaming. On2 VP6 provides higher quality images. On the other hand, this format is more complex, which can make it difficult to view the video files on dated devices. Starting with Flash Player 9 Update 3, the new ISO Base MPEG-4 Part 12 multimedia file format with the new H. 264 video codec is supported. This standard of video compression provides with a significantly more detailed and «clear» image, especially in dynamic scenes.

The AVI (Audio-Video Interleaved) format is a technology designed by Microsoft: the most common and least compressed of all the video file formats. The files created using this method have the.avi extension. Video and audio data are recorded to the same file on the disc in the following way: