Information technology – lecture 9

Representation of computer data.
ASCII and UNICODE.
Text files versus binary files.

Roman Putanowicz
R.Putanowicz@L5.pk.edu.pl

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Different data types

1. numbers
2. text
3. audio
4. images and graphics
5. video

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union
within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

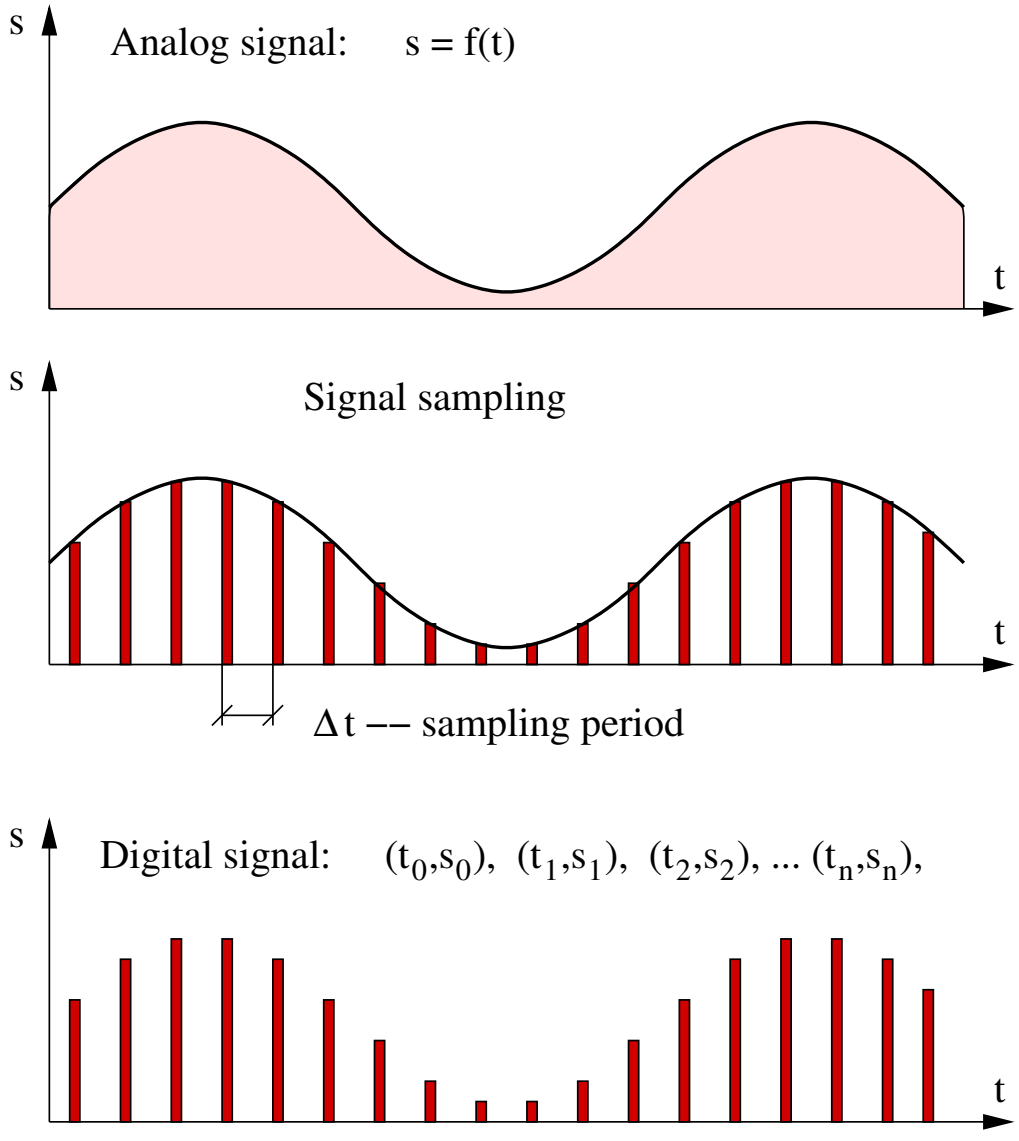
Analog versus digital information

analog information – infinite number of values,

digital information – finite set of values.

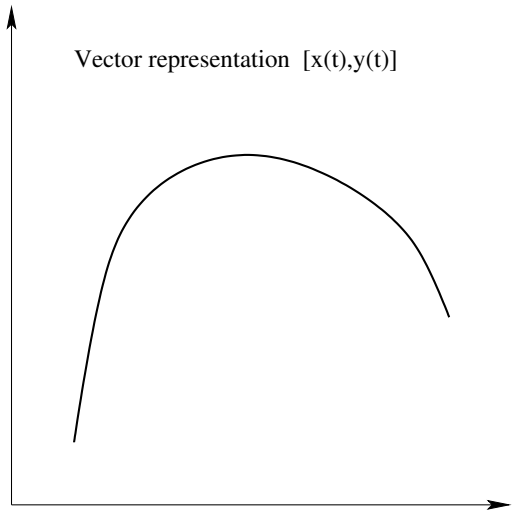
Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

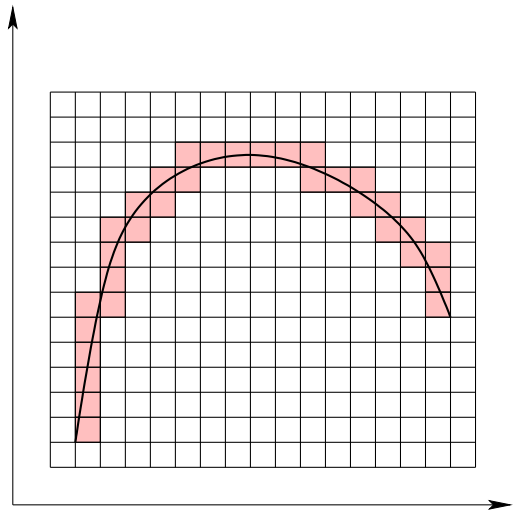
Discretisation example – sampling signals



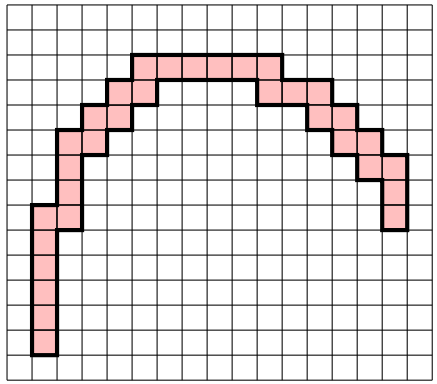
Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

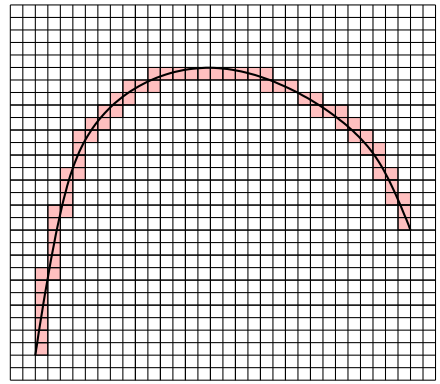
Vector versus raster graphics





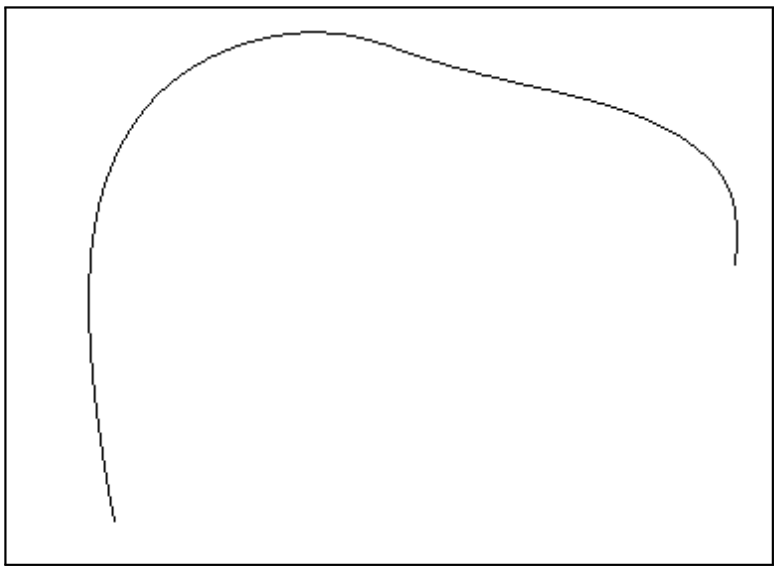
Raster representation

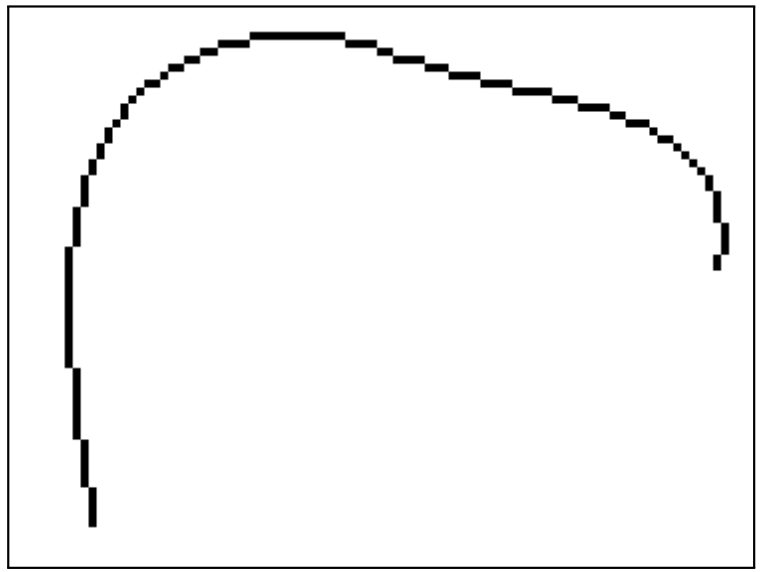




Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Sampling curve at different resolutions





Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union
within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Space utilisation of various graphic file formats

File	Size [B]	Description
Vector formats		
avb.fig	289	FIG image text, version 3.2
avb.svg	1978	SVG Scalable Vector Graphics image
avb.eps	4764	PostScript document text
Raster formats		
avb.gif	997	GIF 358x263 8-bit PseudoClass 2c 997b
avb.png	2496	PNG image data, 358 x 263, 8-bit/color RGB
avb.jpg	4591	JPEG 358x263 8-bit PseudoClass 256c
avb.tif	566136	TIFF image data, little-endian

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Details of FIG format

```
#FIG 3.2 Produced by xfig version 3.2.5b
Landscape
Center
Metric
A4
100.00
Single
-2
1200 2
3 2 0 1 0 7 50 -1 -1 0.000 0 0 0 4
 765 4185 1170 1035 4365 1215 5220 2340
 0.000 -1.000 -1.000 0.000
```

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Data compression

By data compression we understand reducing the amount of space needed to store a piece of data.

lossless compression – class of data compression algorithms that allow to store data without loss of information. Lossless graphic formats: PNG, TIFF.

lossy compression – by approximating the original data some information is lost but in exchange for better compression rates. Lossy graphic formats: GIF, JPEG.

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Binary representation of data

Computers can only manipulate information that is encoded in a sequence of bits of a finite length.

bit – basic information unit, the amount of information that can be stored by a digital device having only two distinct states.

byte – ordered sequence of bits (usually 8)

word – a unit of data specific for a particular computer architecture

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Numbers

Number categories

- natural numbers
- integer numbers
- rational numbers
- real numbers

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union
within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Positional notation

Let:
$\beta \in N, \beta >= 2$ – the base
x_k – digits, $0 <= x_k < \beta$ with $k = -m, \ldots, n$

Notation:

$$x_\beta = (-1)^s [x_n x_{n-1} \ldots x_1 x_0 \, . \, x_{-1} x_{-2} \ldots x_{-m}] \quad x_n \neq 0$$

Interpretation:

$$x_\beta = (-1)^s \left(\sum_{k=-m}^{k=n} x_k \beta^k \right)$$

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union
within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Endianness

The way of ordering individually addressable sub-units, that is words, bytes, bits. In most cases endianness refers to the order of bytes within a word. Two most common ways:

little-endian – increasing significance with increasing memory address

big-endian – opposite, most significant byte first.

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Representing text

Character set – list of characters and the codes to represent them.
The two most popular standards:

- ASCII
- Unicode

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

ASCII

American Standard Code for Information Interchange (ASCII) – originally 7 bit character set (128 characters). The extended version (8 bit character set) called extended ASCII (high ASCII, Latin-1 Extended ASCII) allows to include special characters, for instance accented letters, thus allowing to handle character sets for languages other than English.

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Unicode

A standard for encoding, representation and handling of text, that includes most of the world's languages. Unicode defines various encodings:

- UTF-8 – a 8 bit, variable width encoding (ASCII compatible), uses 1 to 4 octets (bytes) for each character.
- UTF-16 – a 16 bit, variable width encoding, uses one or two 16-bit code units.
- UTF-32 – a 32 bit, fixed-width encoding.

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education

Plain text files versus binary files

plain text file – the contents of the file is readable as textual material without much processing

binary file – the content of the file must be read by a program that interprets it according to some specified binary file format. When opened in a text editor one gets usually unintelligible display of textual characters.

Project "The development of the didactic potential of Cracow University of Technology in the range of modern construction" is co-financed by the European Union within the confines of the European Social Fund and realized under surveillance of Ministry of Science and Higher Education