Unit 1. Information First !

Objectives of the unit

Upon completion of the unit the student should know about

- Information and meaning
- Bits and binary codewords
- Analog and digital techniques
- Methods for encoding
- Redundancy
- Compression
- Encryption

Introduction

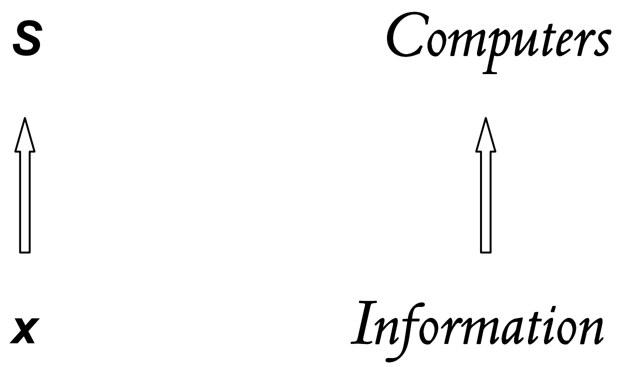


Figure 1-3

The shape of the system S that manufactures the product x comes before x in the world. Conversely tuition goes forward in the inverse order: the study of x introduces S. There are several examples of this systematic approach. E.g. Specialists take lessons on chemistry, organic chemistry, physics before tackling the plant S that refines raw the petroleum x. They consider the refinery processes only when they are familiar with oil and derivatives.

Nobody has ever challenged this natural method, necessary for the acquisition of complete knowledge. We run the same way in informatics.

Computers manipulate information and we firstly shall revolve around this essential element, then we shall be able to mature the hardware and software solutions.

Linguistic Remark: Philosophers make subtle distinctions between the terms; instead we take *Computing, Computer Science, Informatics* and *Information and Communication Technology (ICT)* as synonyms from now onward.

Information is Physical

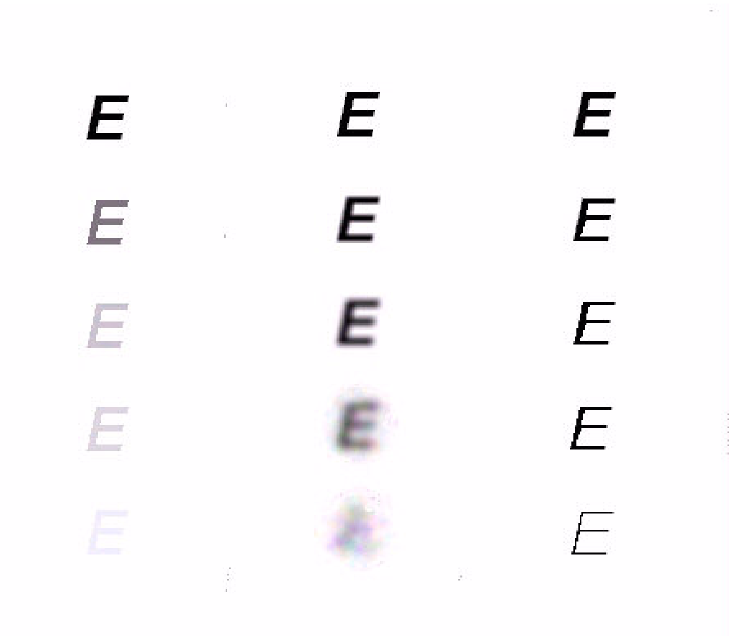


Figure 1-4

We daily exchange several forms of information: pieces of news, photos, songs, gossip and numeric values are common examples. I could remind many others as a large amount of messages floods on us. More or less truthful notices, more or less interesting voices, more or less important signals bombard our lives.

Generically speaking we are familiar with information. However, your naïf approach provides a generic conception that mismatches professional responsibilities. You should obtain the precise definition of information through mathematical calculation, but this direct way appears demanding and I select a more plain sailing.

Take a printed letter over this page. The letter is a piece of information when it contrasts with the white sheet. If the letter is not distinct because it has a blurred outline, or it bleaches, or it thins out etc., information disappears. Facts tell us that an item of information must be discernible and that, missing this characteristic, it vanishes. Consequently, the entity E is said to be information if it differs from the close item E*

$$\mathbf{E \neq E^*}$$

Television news, journals, books, technical reports, films are pieces of information due to the property of being distinct. If they blur or become fuzzy, they are no longer information. A piece of information is a substance that is sharp and neat to comply with the above inequality.

People, devices, appliances manufacture messages which require liters of ink, tons of paper, Kwatts of energy. These physical quantities uphold information is a substance. The inequality sums up endless experiences in the field and qualifies the material nature of information.

Note that technologies deal with products and this rule is necessarily true even in our field. If you suspect information is ethereal, than you waste money because machines manipulate nil...
An evident contradiction.

The abstract and incorrect interpretation of information is capable of diverting your mind and impairing your intellectual efforts for life. As the incorrect attitude of a child may subvert his whole lifetime, similarly a wrong notion in the first stages is able to endanger your whole professional preparation on informatics.

Bits and Binary Words

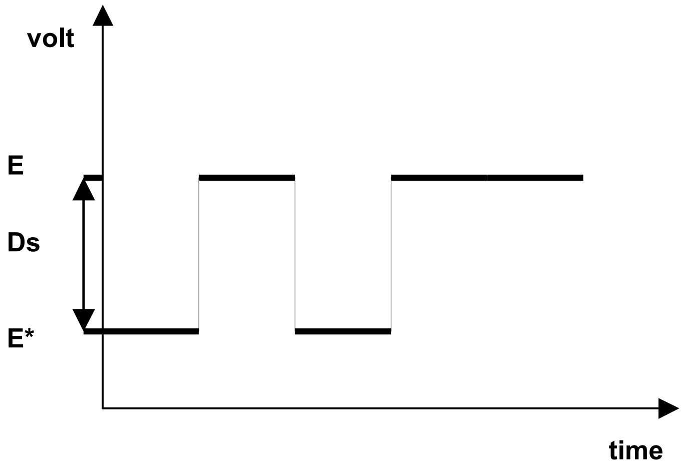


Figure 1-5

ICT applies the inequality **E ≠ E*** with absolute precision and we first examine the most accurate pieces in use: the bits and the binary words.

The bits E and E* are electrical impulses that run along the wires in the computer circuits (Figure 1-5). They are magnetic spots in the floppy and the hard disks; they are coupled mechanical levers in mechanical devices; they are pits and solid surface in the compact disk; they are flashes of light and dark spots in the optical fibers. Bits are physical entities, notably they are two distant values whose **separation Ds** specifies how much they differ. For ease the bits in Fig 1-5 have the following voltages

E = 1.0 volt **E* = 4.5 volt**

Ds = (4.5 - 1.0) = 3.5 volt

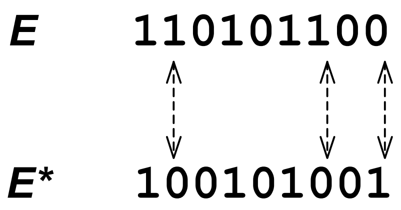
The separation qualifies the ability of the bits to inform. In fact the more they are distant and the more they cannot be confused and can be perceived under the worst conditions. When noise and attenuation put them nearer, bits are distinguishable nevertheless, due to the separation Ds.

Linguistic Remark: Technical literature usually denotes E and E* with 1 and 0. These figures make evident the bits are distinct

$$\mathbf{1 \neq 0}$$

Unfortunately, the symbols 0/1 hint the idea that computer systems process abstract numbers. The term "bit" contracts the words “binary digit" and reinforces wrong suspicions in the mind of people. They presume the calculator machines process immaterial entities instead an appliance can manipulate but material items by definition. Any piece of information, a bit as well, is physical. Computers do nothing but receive and transform material items. We have to contrast with any abstract idea that jeopardizes the correct interpretation of machines.

Bit and Binary Words (contd.)



E 110101100

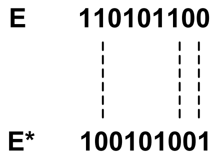
*E** 100101001

110101100 ≠ 100101001

***d* = 3**

Figure 1-6

Information is a substance, hence engineers combine 1 and 0, and make up an informational building, namely they prepare binary **codewords.** They work like the mechanic who combines the wheels, the doors, the bonnet and makes a car.
Binary codewords conform to the inequality, the codeword E has at least one bit different from another word E*. For instance, three corresponding bits mismatch hereunder



E 110101100

E* 100101001

Hence they differ and are two items of information

110101100 ≠ 100101001

The number of odd bits, called **distance,** gauges the inequality between the binary codewords E and E*. In the last example

d = 3

Engineers usually design a set of words which have a certain distance one from the other. The minimum value of these distances, named **Hamming's distance,** specifies the overall quality of the encoding. For ease, the minimal distance measures one bit in the ensuing set

00
01
10
11

***minimal d* = 1**

If one bit inverts, a word becomes identical to another and they are no longer distinct. The following code is better distinguishable than the previous and is more reliable in front of risks

0000
0011
1100
1111

***minimal d* = 2**

In fact two bits must change to make a word equivocal.

Natural Information

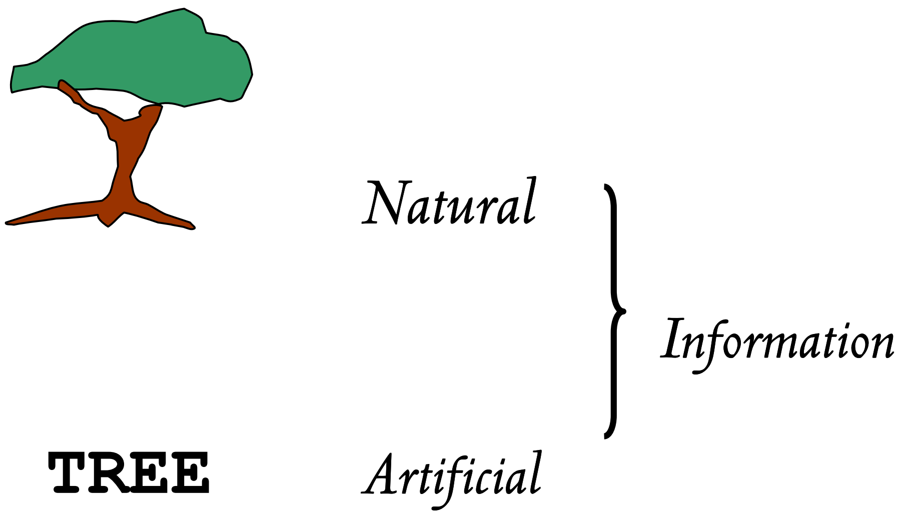


Figure 1-7

If you restrict attention to binary methods, you get a rough and incomplete knowledge of the information and communication technology.

Note how the inequality **E ≠ E*** concerns material entities of any gender, and anything that makes a difference is information. This definition is valid everywhere in the world and expresses the ensuing simple and absolute truth.

The object E, which differs from the close item E*, is information.

Every thing that conforms to this simple rule is information, e.g. the chair, distinct from the floor, makes you acquainted of its existence and is an item of information. Note how this message is essential for you before sitting on it !

Universe diversifies. The mountain Everest, the Sun, a river and a tree are distinguishable hence they are examples of information in the natural state. They are vital to all of us because we cannot survive without them. People can move, eat and work thanks to what they perceive in Nature. Spontaneous information plays an essential role.

The complete set of information includes free and **natural** pieces, like the tree and the chair, and **artificial** items, like printed words and electric bits, that men/women generate for the special purpose to inform. This criterion for grouping is not new and other disciplines share the same approach. For example, chemists study the substances in the natural state and in laboratory.

Artificial information relies upon technology and man accomplishes the pieces by means of tools, instruments and machines, but there are also messages prepared without any technology. Humans and animals are capable of communicating by means of the voice, by the expressions of the face and the body postures, by the emission of smells.

Figure 1-7b

Animals generate informational items, which are alien to any technology and rational criteria. Take for instance a quadruped which makes use of its tail to communicate. It keeps the tail between the legs and tells fear. If it wags its tail, it signifies happiness and devotion. The quadruped is antagonistic and aggressive when it raises its tail.

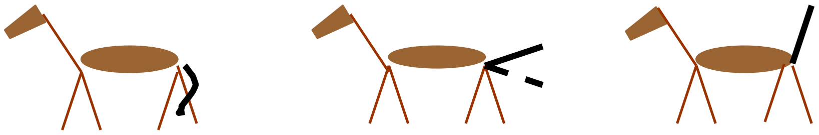


Figure 1-7c

Animals deliver signals without any preliminary educational stage. Men/women exchange instinctive signs as in Figure 1-7b, but technology frequently interferes with human communications. When a man cries, he emits a spontaneous sign of pain, instead when he talks, he encodes his voice using the alphabet that is a digital solution.

Meaning of the Meaning

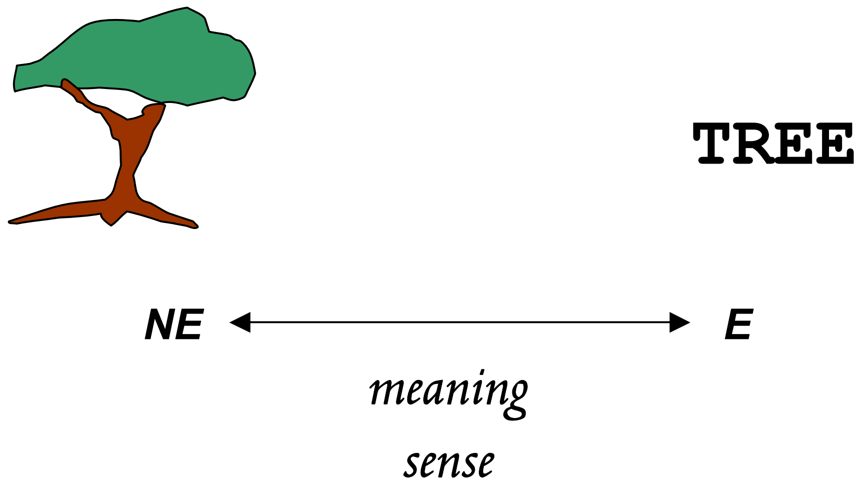


Figure 1-8

Natural pieces of information are vital to mankind and each individual can survive exclusively thanks to what he sees, he listens to, he perceives in the physical world. In short, men/women can live only if they exploit spontaneous information.

This resource unfortunately is bulky or vice versa microscopic. Sometimes it is speedy and does not leave any trace; it is unshakeable in other events. Natural information has so many disadvantages and contemporary is so essential to survive that people build an artifact which substitutes the genuine object. The artificial piece of information E replaces the natural NE, to wit **E stands for NE** and we draw this conclusion.

The basic function of E to represent NE
Constitutes the meaning (or sense) of E.

Take for instance, the town of Rome which you cannot perceive because it is inaccessible from your position. This item of information may be prerequisite but is unavailable and you prepare the word **ROME**, a tiny ink object that you can easily detect, move, manipulate, transmit etc. The word **ROME** is the copy of the physical town, namely the former is the **model** and the latter the **original**.

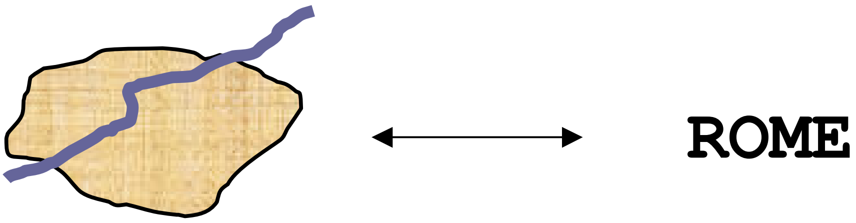


Figure 1-8bis

Meaning is the fundamental property of information and constitutes the core reason for the success of the computer systems, which manipulate, process, transmit and store the models of the reality that are essential to our lives.

Despite the innovative glamour of computers, the birth of ICT goes back more than twenty thousand years. In the prehistory, the graffiti E represented the shooting parties NE that were no longer perceivable. Wives and children, absent from the struggles with the beasts, relived the impressive scenes thanks to the copies of the events.

The discoveries of ICT marked the peoples history. The art of writing, the paper and the printing machine ushered in revolutionary advances of mankind. The number of inventions has increased over the last couple of centuries. We actually purchase new equipment at an accelerated rate and machines are invading our lives. Fax and hi-fi, television and radar, mobile phone and computers have broad influence on our personal relations. Since prehistory, man is still in the ICT way due to the vital importance of natural information NE and, in turn, of the copies E.

Meaning of the Meaning (contd.)

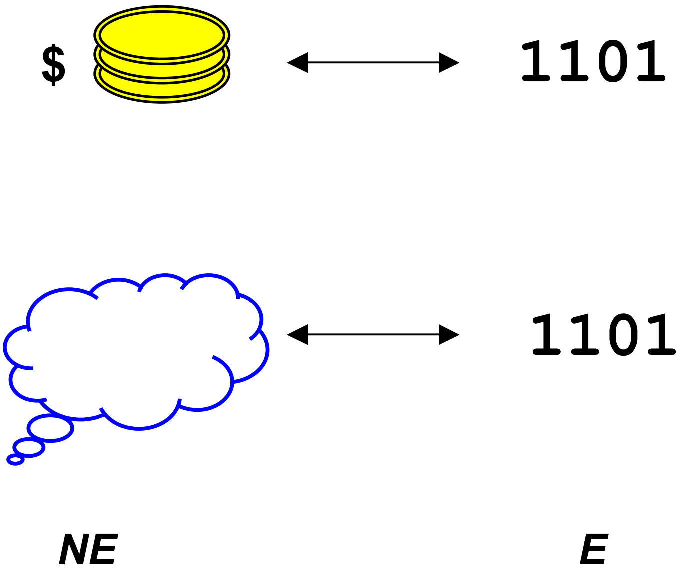


Figure 1-9

Information is a substance and E stands for a practical entity. The artificial item E is the counterpart of the spontaneous item NE. This rule does not exclude exceptions and we pay attention to information E representing an abstract entity instead of a material one. The "number" NE that is something mental and is denoted by the numeric word E, provides a valuable case. We spot these special cases and put apart the numeric word representing an abstract entity from the numeric word representing a material entity such as dollars, a machine, an animal etc. We discriminate "3" that means three eggs, from "3" that is a pure number. We recognize "1101" denoting the number thirteen and "1101" that marks money.

Concluding, a piece of information has two distinguished peculiarities:

I) **E** is physical,
II) **E** stands for something.

We could say it has a double nature that the following formal expressions summarize with mathematical precision:

I) **$E \neq E^*$**
II) **NE — E**

The former specifies the **physical and technical character of E**, the latter does the **semantic and psychological side of E.**

Such qualifications are essential to assume a correct skill in informatics. Instead several individuals confuse E with NE, they merge the physical value NE with the abstract number NE and lose their bearings in the field.

PROS & CONS

NATURAL INFORMATION:

- Vital to survive
- Bulky, volatile, difficult to detect etc.

ARTIFICIAL INFORMATION:

- Trivial in itself
- Essential substitute of natural item.

Pathway to Success

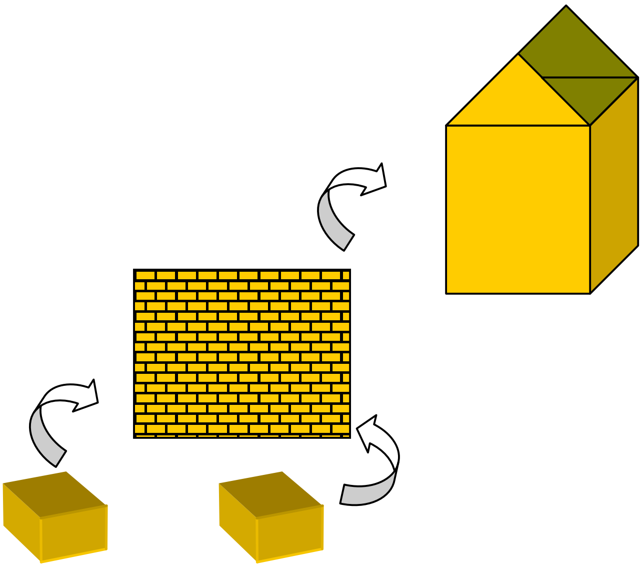


Figure 1-10

We have discovered that information is physical, hence ICT experts are able to construct information like a building. Notably they assume a set of basic elements and produce a vivid message through combinatory methods. People introduced this procedure when they invented the alphabet thousand years ago; scientist perfected it in the modern age.
The **digital technology** constructs out information by a progressive method and achieves all the informational forms through only one rule. This standard is the key to its overwhelming success, as we shall see in detail.

I sum up the preparation of digital information in five principal steps.

1° Hardware engineers establish the elementary information items: the **bits 1** and **0**.

2° They prepare the **binary word** which stands for a character, or a figure, a symbol, a sound, a color etc.

3° Experts or users combine common words, numbers and other simple structures by the mean of binary words. E.g. The name **ROME** consists of the ensuing symbols which in turn are made by bits

R	**O**	**M**	**E**
01010010	**01001111**	**01001101**	**01000101**

We build the number **37** as such

3 7

00110011 00110111

4° Users prepare a **text**, a **document**, a **picture,** a **piece of music** etc. which are complex structures by combining the previous components.
5° At last they unite the variety of informational forms coming from the step 4° and obtain a **hypermedia**.

The digital technology is rigid and precise in the first two stages which pertain to hardware specialists. The items of level ***3°*** are designed through the precise criteria, which we shall see next, or otherwise may be created by means of imagination. The items at levels ***4°*** and ***5°*** are always arranged under psychological and artistic emotion. People set them up according to communicative feeling. In conclusion, the outcomes of the last stages are rich in contents and figment; instead the first pieces of information derive from calculations.

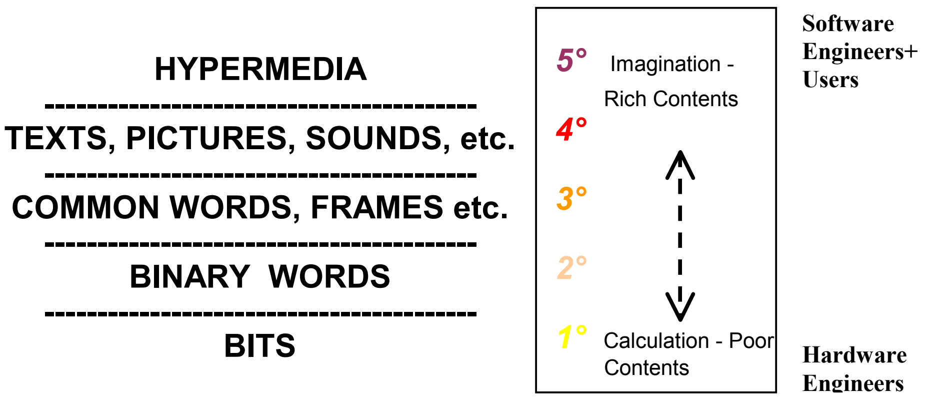

HYPERMEDIA
TEXTS, PICTURES, SOUNDS, etc.
COMMON WORDS, FRAMES etc.
BINARY WORDS
BITS
5°
Imagination - Rich Contents
4°
3°
2°
1°
Calculation - Poor Contents
Software Engineers+ Users
Hardware Engineers

Figure 1-10bis

Digital manufacturing looks like the building of a house. Bits ensure the perfect quality of the informational construction likewise the bricks that pledge the regular form of walls and rooms. The largest parts of the house are designed under human tastes so levels ***4°*** and ***5°*** rely on people communication requirements.

Linguistic Remark: The international symbols 0 and 1 represent the bits and firmly state this inequality is true

$$\mathbf{1 \neq 0}$$

Pay attention. The figures 0/1 seem to suggest the bits are complex whereas they are ***elementary.*** This feature is essential to the gradual combination process of Figure 1-10 which otherwise would appear unreliable and untidy. If you mix up the ciphers 0 and 1 at level ***1°***, with those at levels ***2°*** and ***3°***, the digital approach becomes confused and you cannot grasp the perfect methods pertaining to technology.

Pathway to Success (contd.)

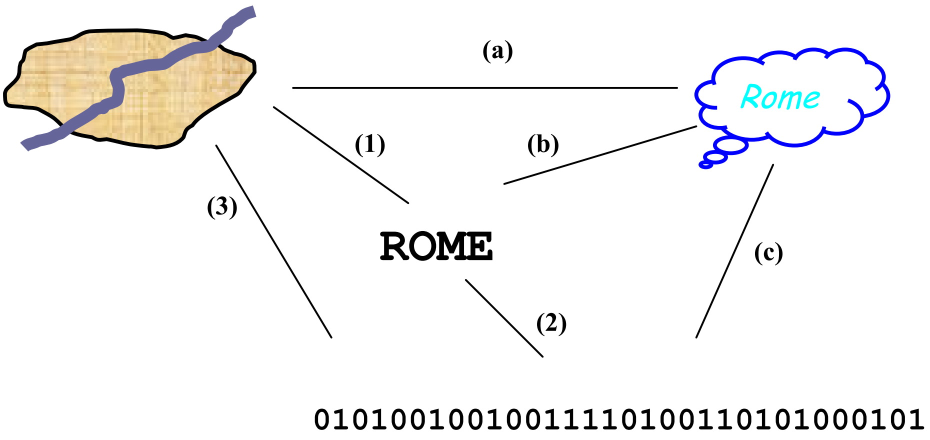


Figure 1-11

Circuits handle bits, but humans slog through binary words and go for the common alphabet. This choice generates an intricate net of copies and significance.
For example, the word ROME stands for 01010010010011110100110101000101 (=2) and in turn marks the physical town (=1). Moreover humans have the idea of Rome in the brain. This mental information regards the physical town (=a), and is even denoted by both the digital items (=b and c).

Digital information has intricate semantic relations indeed!

Specialists in computer science should be familiar with semantics instead the modern "theory of information" overlooks this argument and curricula frequently omit lessons on semantics. Engineers narrow their interest to the relationship (2) because of this insufficient education, and neglect the other meanings that compound their works. Partial studies have deep roots and this book aims at offering a valuable aid to bridge the gap.

* * *

The method in Figure 1-10bis entails that we are able to create any information by means of the suitable string of bits. Nothing confines the possibilities of construction on condition that the progressive assemblage is correct.
Moreover we can produce whatever information by combining a given set of bits in input with the appropriate rule. We shall comment this extraordinary quality of the digital technology in the next chapter.

Encoding

Digital **B > 2**

Binary **B = 2**

Figure 1-12

Four steps out of five in Figure 1-10bis manipulate codewords and I provide explanations about the encoding criteria that impact on both the computer hardware and software.

Encoding consists in building several codewords by means of combining the symbols of a prearranged set.

In particular **binary encoding** works with only two elements at level ***2°***

B = 2

The adjective **digital** marks the encoding at the stage ***3°*** that works with a variety of bases such as the *Arabic figures* E.g. Figure 1-16.

B = 10

The international *Latin alphabet* has twenty-six letters

B = 26

The *alphanumeric base* groups the Arabic figures and the alphabet (E.g. Figure 1-17)

B = 10 + 26 = 36

Now you should be capable of distinguishing whether

10110

Is the binary word at level ***2°*** with the base

B = 2

Or otherwise it is the numeric word at level ***3°***

B > 2

Encoding (contd.)

$$N = B^L$$

$$N = 10^2 = 100$$

`00 01 02 03 04 05 ..... 98 99`

Figure 1-13

Technologies require computations and measurements, binary and digital techniques do the same.

The number **B** of basic symbols and the length **L** of the codewords are the initial parameters for encoding. Specialists measure these quantities in the following units

- **bit** when the base is binary,
- **byte** (or **character)** in the other cases.

As an example, we say

`101` is 3 *bits* long

`5488` occupies 4 *bytes*.

The measurements of volume apply to the container and the content, and this rule is valid in ICT too. The length gauges the codewords and also the stores, the memories etc. which contain them. Storage is measured by the byte multiples because of its extension

Kbytes	~	thousand bytes
Mbytes	~	milion bytes
Gbytes	~	billion bytes

The basic formula for coders yields the **number of words** from the length and the base

$$N = B^{L}$$

As an example, the binary words with 8 bits are two hundred and fifty six

$$N = 2^{8} = 256$$

The total amount of codewords from 00 to 99 is

$$N = 10^{2} = 100$$

From the alphabetic keys we obtain 529 words of two characters

$$N = 23^{2} = 529$$

Encoding (contd.)

ASCII Code

0100 0000 → **A**
0100 0001 → **B**
0100 0010 → **C**
.... ..

Figure 1-14

Computer systems manipulate, store and transmit binary data. We remind the **general-purpose codes**

- American Standard Code for Information Interchange (ASCII)
- Extended Binary Coded Decimal Interchange Code (EBCDIC)

They are international standards. For example, networks exchange signals coded in ASCII, programs of mainframes handle data in EBCDIC.

Computer systems handle **special-purpose codes** too, notably most peripherals work with special binary codes. They depend on subtle technical considerations which I cannot comment on. I restrict to an example in the next pages.

Encoding (contd.)

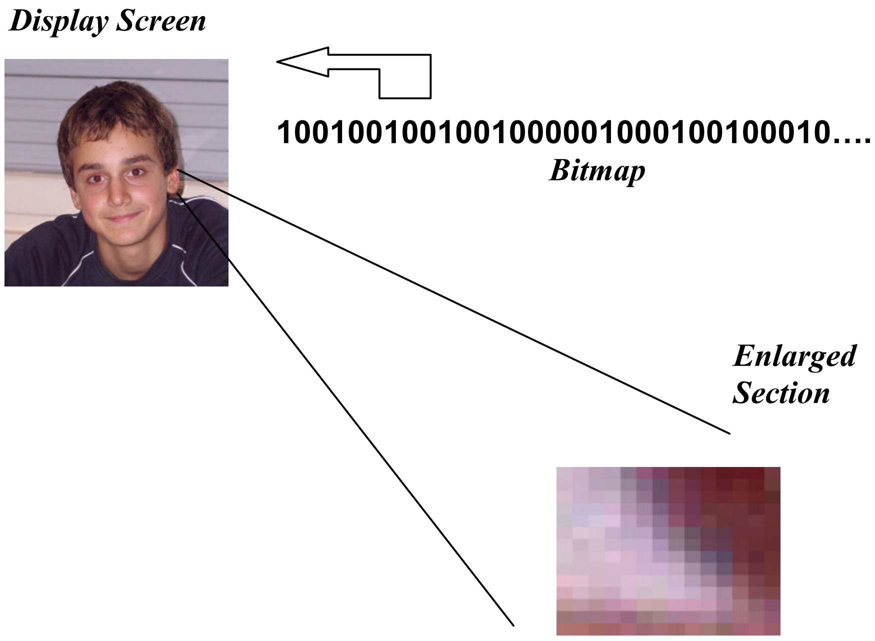


Figure 1-15

The screen-device displays a set of minute picture-points (shorten **pixels**) invisible to unaided eye. People can distinguish a pixel only if they use common lens or enlarge a section of the image.

The display screen is necessarily supported by the **bitmap,** a matrix of binary codewords. In fact, each codeword of the bitmap assigns the color to a pixel. If the words are one bit long, the pixels may be black or white

$$\mathbf{N = 2^1 = 2}$$

The Video Graphics Array (VGA) displays 16 or 256 colors. The codeword is four or eight bits log, namely it covers half-byte or one byte

$$\mathbf{N = 2^4 = 16 \qquad N = 2^8 = 256}$$

From 1990, the Extended Graphics Array (XGA) offers 65 thousand or 16 million colors, namely 16 or 24 bits (respectively 2 or 3 bytes) encode a pixel

$$N = 2^{16} = 65{,}536 \qquad\qquad N = 2^{24} = 16{,}777{,}216$$

How a binary codeword symbolize a pixel ?
Usual monitors work with the Red-Green-Blue (RGB) system, to wit the specific color of a pixel is some blend of three pure colors. RGB determines 24-bit encoding that devotes eight bits to each basic color and generates the most ample palette with 16,777,216 tones of colors

$$N = 2^{24} = 2^{8+8+8} = 256 * 256 * 256 = 16{,}777{,}216$$

The value 11111111 stands for the pure color, 00000000 means the color is suppressed. To exemplify, green, red and blue are obtained by the following codewords

111111110000000000000000 = **green** (red and blue suppressed)
000000001111111100000000 = **red**
000000000000000011111111 = **blue**

Intermediate binary values bring the spectrum of hues

000000000000000001000000 = **navy blue**
000000000000000000100000 = **dark blue**

The thorough combination of red, green and blue gives a white pixel. When they are absent, the color is black.

111111111111111111111111 = **white**
000000000000000000000000 = **black**

The proportion of green, red and blue brings the variety of tones

111111111111111100000000 = **yellow**
111111110000000011111111 = **fuchsia**
000000001111111111111111 = **aqua**

Usual Internet browsers interpret 6 values of each basic color, they display 216 hues instead of 16 millions

$$N = 6 * 6 * 6 = 216$$

When an image has more tints than the program can display, the browser simplifies the visual rendering.

Encoding (contd.)

α

010	*Switch*
015	*Mower*
020	*Lamp*
025	*Folder*
030	*Shears*
035	*Lathe*
040	*Transformer*

β

101	*Folder*
102	*Shears*
103	*Lathe*
...	....
201	*Mower*
...	....
901	*Transformer*
902	*Switch*
903	*Lamp*

Figure 1-16

Software specialists and users as well work at step ***3°***, they design common words or otherwise invent new codewords. They are rarely guided by imagination, they prefer a solution according to the ensuing criteria.

a) – Broadly speaking, the natural language inspires the clearest codewords.

* As an example the *country code* consists of two letters copied from the country name

DE = Deutchland
FR = France

IT	= Italia
NE	= Netherland

The airlines apply this same criterion

AF	= Air France
AI	= Air India
AM	= Aeromexico
AR	= Areolineas Argentinas
AT	= Royal Air Maroc
BA	= British Airways

They are said **mnemonic codes** because they are very easy to remember**.** As a counter-example the codes in Figure 1-16 are not mnemonic.

b) - The **telling code** expresses the categories of the objects, of the individuals, etc.

* As an example the first cipher of the group β (Figure 1-16) "tells" the category of machines.

* The following page (Figure 1-17) lists the codewords whose first figure highlights the building class. We can distinguish the houses hosting individuals, communities etc.

c) - Software programs need ever new information and specialists frequently insert, change or cancel the codewords. Codes are to be flexible. The **discrete code,** which leaves spaces among the words, provides a solution for updates. The codewords are abundant and longer then the codewords strictly necessary to symbolize the original objects.

* Both the codes in Figure 1-16 may encompass new numbers (omitted in the list) which may be assigned to new categories.

Encoding (contd.)

Buildings

A1	***High-class houses***
A2	***Civic houses***
B1	***Colleges***
B2	***Hospitals***
B4	***Public Offices***
B5	***Schools***
C1	***Shops***
C2	***Stores***
C3	***Deposits***
C4	***Hangars***

$$E \longleftrightarrow NE$$

Figure 1-17

Software specialists usually design transparent, flexible and secure codes to achieve high quality information: **data**. I set apart data that are established through the **coding plan** from generic and ambiguous pieces of information.

The plan directly derives from the formula **NE—E**. In fact this expression has two terms that generate the two-entry table: one column shows the values E and the other column lists NE.

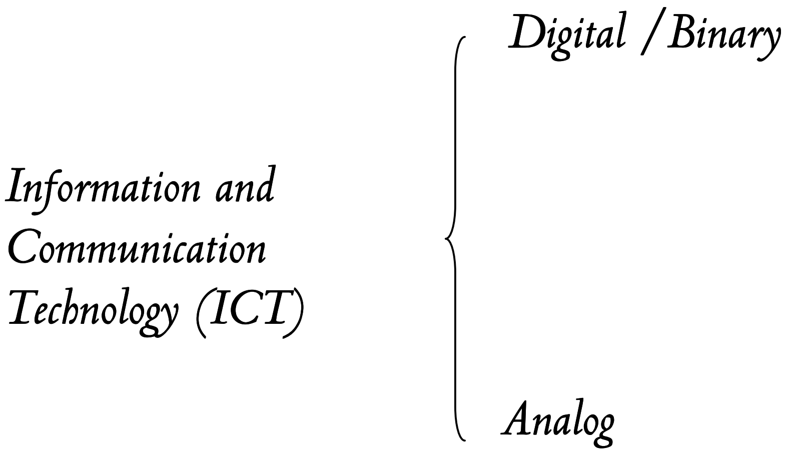
PROS & CONS

ENCODING:

- Mnemonic and telling codes are transparent
- Discrete codes are flexible

- Codewords longer than necessary need large spaces in storage and delay transmissions

Analog Technology



Digital / Binary

Information and Communication Technology (ICT)

Analog

Figure 1-18

Digital technology applies this inequality **E ≠ E*** along the progressive construction of information which goes on through one rigid track.

In the past, man was ignorant of mathematics and did not explore precise approaches. He simply shaped artificial information by copying spontaneous items. A piece that imitates Nature is "analogous" with a natural entity, namely it is **analog.**

Engineers call *analog* all the pieces of information that do not comply with the digital method. They are still in use because easy and economic. Do not underestimate the analog technology that provides fruitful and notable solutions and constitutes the second main field in ICT.

Digital technology steers only one course. Every informational form is constructed through the steps ***1°, 2°, 3°, 4°*** and ***5°***. Conversely, analog technology does not lay down any standard procedure. Engineers, inspired by natural objects, by a physical law or a chemist property invent a solution whenever it serves. For example:

- Edison registered the voice in a waxed cylinder.
- Meucci invented the telephone exploiting the properties of coil.
- The photocell is the chief resource for cinema sound track.
- The tape-recorder utilizes the qualities of the magnetic paste.

Analog Technology (contd.)

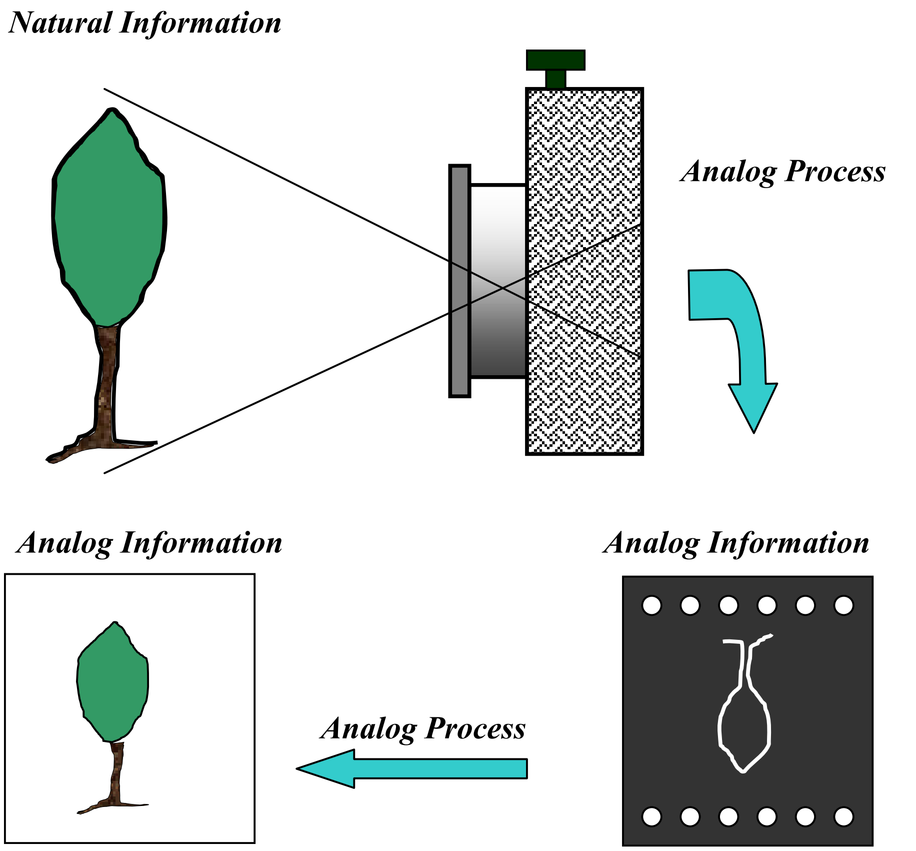


Figure 1-19

Analog devices require an extensive study because each solution is original and unique. This topic goes beyond the scopes of the present volume. I barely bring an example to illustrate how analog technology can “imitate Nature”.

The camera has the film at the back and a chemist reaction forms the image over the film. The natural object directly determines the negative picture that has no bit or binary words, and appears analogue to the natural image.

Chemical reactions achieve the final picture on the paper from the negative. Also this result constitutes an analog piece of information.

Note how the analog processes are not progressive, they do not use small pieces to achieve an ample outcome.

Some scholar claims: “Analog technology is continuous”.
This statement is true. Natural signals (e.g. music, sounds etc.) are continuous during the time and analog devices convert them into the uninterrupted symmetrical form.
For ease, the microphone translates the voice that is a continuous wave into the electrical signal that has an identical shape.

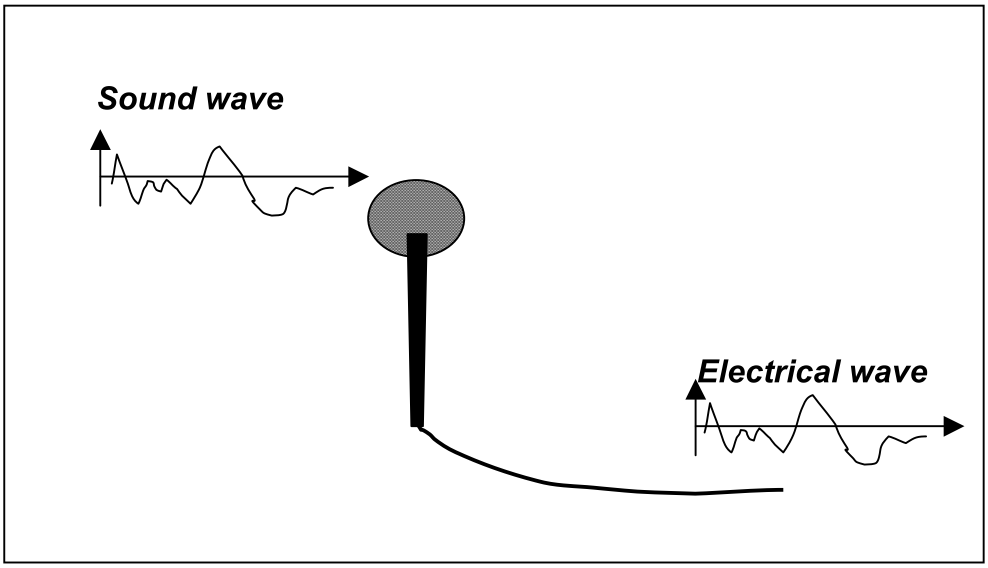


Figure 1-19bis

Analog Technology (contd.)

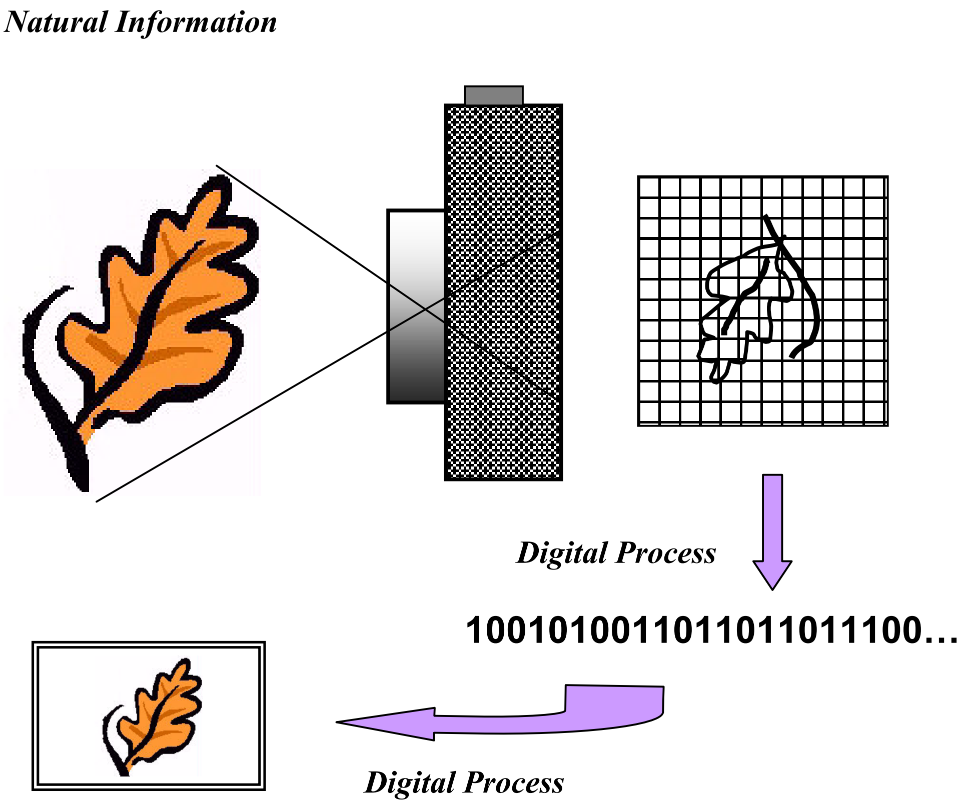


Figure 1-20

The contrast with digital technology reinforces your understanding of analog methods.

Take the modern **digital camera** which includes a matrix with semiconductors on the back. This matrix converts the image of the leaf into a sequence of bits, namely it makes a bitmap including binary codewords. This method complies with the steps in Fig 1-10bis.
Another digital equipment is necessary to interpret this binary data set and to display the definitive outcome to people. Usually the computer monitor converts the bitmap into the image of the leaf through digital processes.

Concluding, digital and analog technologies proceed through rather opposite methods:

- The latter is empiric and occasional.
- The former is systematic and meets with universal success.

Digital technology tackles each application with a standard approach. It even surmounts obstacles and critical events with this style.

PROS & CONS

ANALOG INFORMATION:

- Vivid and immediate as natural information
- Any solution is original and techniques are empirical
- It follows a large variety of approaches.

DIGITAL INFORMATION:

- Cool construct
- Only one rationale technique
- Systematic as it follows a rigid progression.

Unique Strategy against Emergencies

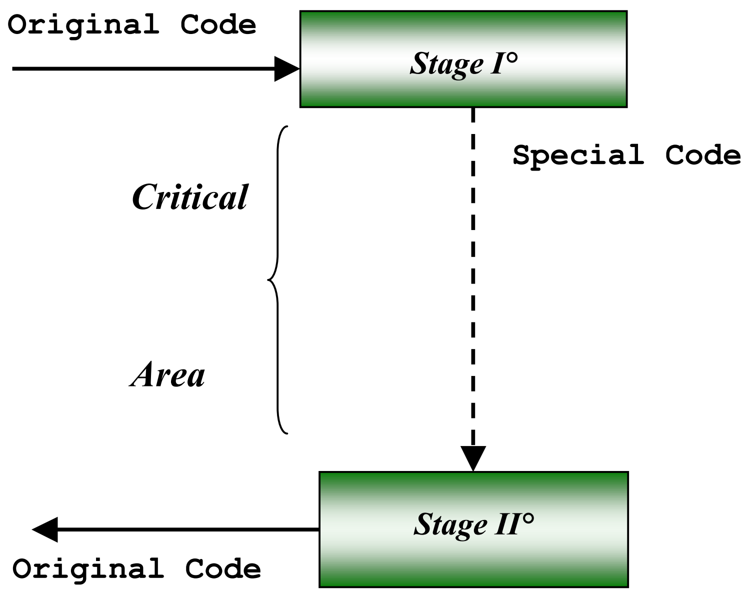


Figure 1-21

Digital technology solves even critical situations through a general scheme. In detail it performs the following operations:

Stage I° - A special code takes the place of the regular code before possible crisis and risks,

Stage II° - The initial code is restored at the end of the hazardous range.

This general implementation confirms the remarkable value of digital standard. It supports universal success whereas analog technologies need to invent ever-new solutions.
Which risks do engineers take?

Information is physical and this quality entails cumbersome volumes, large occupation of space, transport delays, the threat of thefts. In short, information meets the same dangers as a product does when it is stored and moved. Engineers tackle three usual problems:

(A) - How to reduce the volumes of data;
(B) - How to make data more robust;
(C) - How to shelter data from robberies.

(A) - Compression - Electronic and magnetic archives becomes ever more large; notably pictures, movies and songs occupy large memories that increase the transmission delay times and require a large amount of space. In I° technicians reduce the size of the files to handle them more easily. In the stage II°, the files return to the original form for human uses. **Compression** and **decompression** processes are lossless or otherwise lossy.

Lossless – Data exactly return to the initial configuration. There are various algorithms to do this. For instance, the compression stage removes the string of repeated characters and replaces them with a single character and the string length

$$\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat\flat \Rightarrow 23\flat$$

A second technique substitutes smaller bit strings for frequently occurring characters. Take this example

111100010100 111100010100 ⇒ 10 10

Lossless compression provides perfect outcomes that are identical to the input stream. This service is necessary for telecommunications, for programming, for handling documents. Compressed files have usually zip, gzip, winzip or other formats, and may be reduced up to 50% of its original size.

Lossy – This method does not restore information to its precise form in point II°. Lossy techniques are generally suitable for video and sound, where most people do not detect a certain amount of information loss. Lossy decompression does not reconstruct all the bits of input but this shortage is usually imperceptible. You frequently can select the degree of approximation. For example, the jpg file stores an image that has lossy compression. When you create a jpg or convert an image from another format to a jpg, you are asked to specify the quality of image you want. Since the highest quality results in the largest file, you can make a trade-off between image quality and file size. Lossy techniques provide high reduction of space, e.g. 5:1, 6:1 or more.

PROS & CONS

LOSSLESS COMPRESSION:

- All the data kept
- Small compression rate.

LOSSY COMPRESSION:

- Details lost
- High compression rate.

Unique Strategy against Emergencies (contd.)

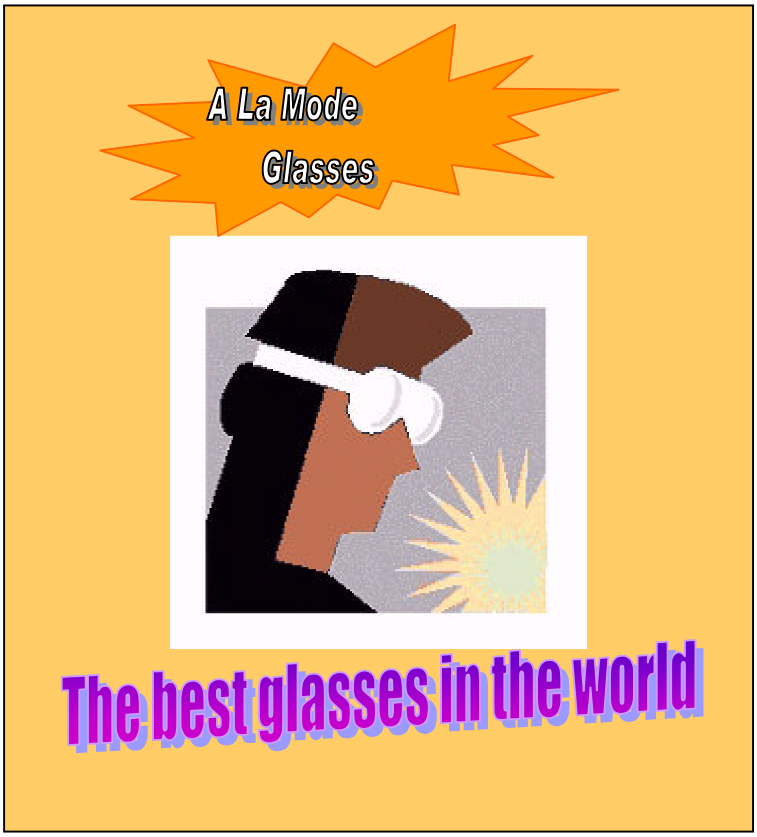


Figure 1-22

(B) - Redundancy - Dictionaries define *"redundant"* something abundant with respect to its usage. Repeated messages are redundant since they represent several times the same content. For example, **ROME** is sufficient to tell the physical town and the double word **ROMEROME** is lavish.

The difference between **n** pieces of information **E** and **m** illustrated items **NE** quantifies the redundancy

$$U = n - m \geq 0$$

E.g. the poster in Figure 1-21 promotes the glasses. The banner, the picture and the final slogan substantially deliver the same message

$$n = 3$$
$$m = 1$$

$$U = 3 - 1 = 2$$

This value qualifies how much the poster is redundant. Conversely redundancy is null and information *optimal* when **n** pieces are just enough to figure **m** contents

$$n = m$$
$$U = n - n = 0$$

Stressing modern life makes harder to listen to. Mass media distract the audience and a communicator is unable to catch the attention of people. A sole piece of news may be ineffective and duplicated pieces are more persuasive than only one. They hold the spotlight even if the hearer is tired and inattentive. Redundancy improves the semantic quality of communication and reinforces its physical nature too.
Redundant messages are more robust in front of possible attacks. A repeated message has double probability of surpassing a dangerous range. They become necessary when noise, perturbations and failures threaten communications.

In substance, critical factors worsen the reception of information both from the semantic and physical viewpoints. Redundancy achieve suitable countermeasures

- As it facilitates comprehension and attention of people,
- As it ensures endurance in front of risks.

Psychologists, humanists etc. focus on the first advantages. Engineers fulfill the second duty by a rational approach.

Unique Strategy against Emergencies (contd.)

a

01111001	1
10000001	0
10110101	1

b

873214
8+7+3+2+1+4 = 25

⇓

87321425

c

12	65	24	91	65		257
24	33	76	66	34		233
94	44	87	11	39		275
						⇓
130	142	187	168	138	⇒	765

Figure 1-23

Machines and software programs make the messages redundant in the stage I° (Figure 1-21). Thanks to this method, they prevent disturbance, noise and random irregularities before hazards occur. Repeated pieces of news increases the probability that the receiver gets one correct message at least.

Duplicate messages although increase the costs; moreover they delay transmission and need large memories. In order to reduce this negative impact, technicians do not duplicate systematically the words but accomplish moderate forms of redundancy. They tune redundancy to the needs and make the codeword E somewhat longer than the size sufficient to represent NE. The implementation observes precise rules so that the receiver in the point II° carries on accurate controls and can take immediate corrective actions in case of errors. I bring three classical examples to illustrate the rigorous checks accomplished by digital appliances.

a One bit is added to the binary word in stage I°. This bit (see the last bit on the right) is switched so that the bits 1 are even in the codeword. If the bits are odd after the transmission or storing, the control detects the error. The *parity bit* makes the codeword redundant, as it becomes little longer with respect to the standard.

b The sum of the ciphers is added to the number. The result 87321425 is longer than 873214, which suffices to relate the mathematical value. When the sum is wrong, the control detects the error at stage II°.

c The package of numbers is enriched by a double set of sums. Each number on the right sums the values in the row. The line at the bottom collects the vertical sums. The vertical partial totals and the horizontal partial totals must make the grand total (= 765) in the right corner at the bottom. If they mismatch an error has occurred. The system transmits 24 numbers instead of 15 and is somewhat redundant.
Binary packages are transmitted in networks using similar methods, which today make for sure communications.

These redundant implementations and others are general purposes and confirm the high superiority of the digital engineering.

Unique Strategy against Emergencies (contd.)

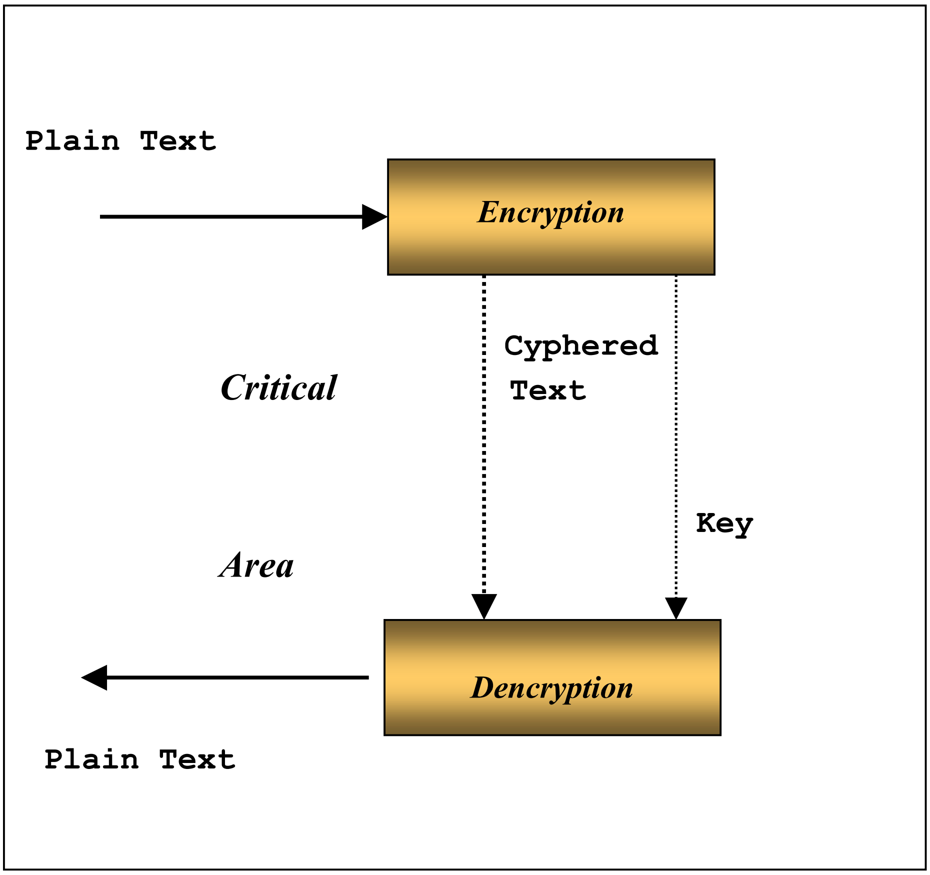


Figure 1-24

(C) - Encryption - Owners protect significant information in the event it may be stolen. They although cannot prevent larceny everywhere and whenever, and resorts to this tactic based on the enormous difference of importance between the physical and semantic side of information. Consider the content of news that is vital, the owner can allow stealing the physical message and forbids unauthorized people to grasp the significance. In fact, the economical value of information is irrelevant because it is tiny. In practice, an algorithmic converts a plain text to a cipher text in stage I°. If an unauthorized person detects the physical message, he cannot comprehend its content. After the critical interval, the text returns to its clear original form, but only the intended recipient of the message can decrypt the message whose meaning is concealed to everybody else thanks to the process in stage 1°.

The words *encryption* and *cryptography* derive from the Greek *kryptos*, meaning *hidden*. In fact, this technique aims at hiding the content but not physical information. The origins of cryptography dates back Julius Caesar, who created a system in which each character in his message was replaced by a character three positions ahead of it in the Latin alphabet.

A B C D E F G H I L M N O P Q R S T U V Z
↓ ↓ ↓ ↓
A B C D E F G **H** I L M N O **P** Q **R** S T **U** V Z A B C

ROME ⇒ URPH

The ability to securely store and transfer sensitive news has proved a critical factor in the Internet and cryptography has turned into a front-line technology. Encryption is the most widely used mechanism for providing confidentiality over an insecure medium.

There are several couples of algorithms to perform the **encryption** and **decryption** processes. The most effective pairs operate through a **key,** a parameter that impacts substantially on the enciphering process with the constraint that the same key must be used for decryption. Frequent changes of the key improve the protection against possible fraudulent deciphering. This 'theoretical' amelioration compounds the problems in the practice. The transmission of the keys to the receiver is critical as the flow of data doubles in the critical area. Reiterated transmission of the keys between the sender and the receiver becomes the weak spot in this secure system.

The **public-keys encryption,** invented in the seventies, overcomes the unsafe management of the keys. This new approach allows the cryptographic keys be frequently changed and openly notified. If a criminal gets the public key, he nevertheless cannot decrypt the ciphered text. In fact, the cryptographic system uses the keys PU and PR with the following features:

- PU is the public and variable key. A user can transmit PU ever again.
- PR is the private and stable key.

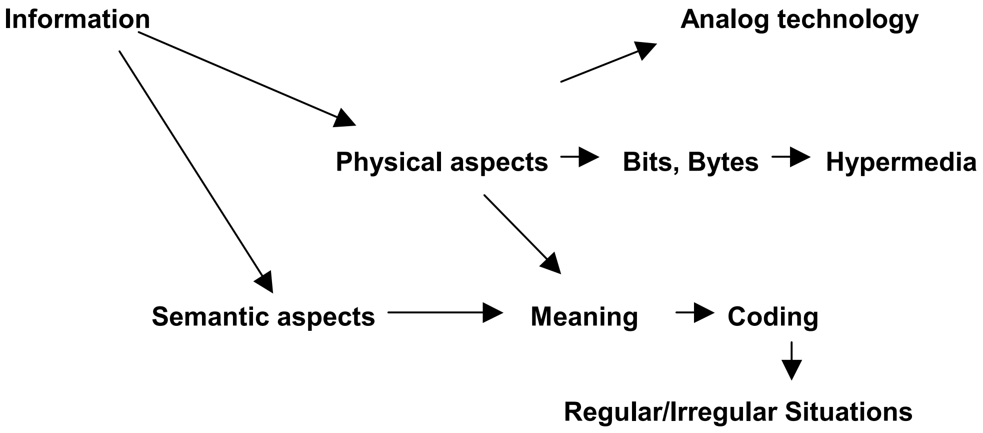
They work as follows. Mr. X encrypts and sends a message to Mr. Y with PU of Mr. Y and only Mr. Y is capable of deciphering the message of X with his own PR.
The pair PR+PU enables the full process, whereas PU is not enough to decipher the message.

Unit 1 SUMMARY

This chapter has clearly examined the physical nature and the semantic capabilities of information. We have discovered the artificial items and the spontaneous pieces of information in Nature.

We have argued the basic properties of digital data from the introductory notion. The stages for the construction have been illustrated from the elementary items up to hypermedia. We have spent pages to introduce the principles of encoding, compression, redundancy and cryptography.

The following conceptual graph sums up the logical path that connects the topics.



This work paves the way toward the unifying view of the information and communication technology that is our final goal.