Representación de la información en un computador.

# REPRESENTACIÓN DIGITAL DE TEXTOS



## Representación de la información en computadores

- L2.0 Sistemas de numeración usuales en informática.
- L2.1 Nociones básicas sobre representación de la información
- L2.2 Representación de textos.
- L2.3 Representación de sonidos.
- L2.4 Representación de imágenes y de video.
- L2.5 Representación de números enteros.
- L2.6 Representación de números reales.
- L2.7 Algoritmos de compresión de datos.

## L2.2 Representación digital de textos.

- Clasificación de los caracteres.
- Códigos normalizados tradicionales: ASCII.
- Unicode.
- UTF-8.
- Consideraciones prácticas.

3

## Clasificación de los caracteres.

- La información se suele introducir en el computador utilizando el lenguaje escrito:
  - Caracteres alfabéticos
  - Caracteres numéricos
  - Caracteres especiales
  - Caracteres geométricos y gráficos (bordes de cuadros, etc.)
  - Caracteres de control

4

- **Caracteres alfabéticos: son las letras mayúsculas y minúsculas del abecedario inglés:**

  **{A, B, C, D, E,..., X ,Y, Z, a, b, c, d,..., x, y, z}**

- **Caracteres numéricos: están constituidos por las diez cifras decimales:**

  **{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}**

5

- **Caracteres especiales**
  - Son los símbolos no incluidos en los grupos anteriores, p.e.:

    **{ ) ( , * / ; : + Ñ ñ = ! ? . " & > # < ] Ç [ SP }**

    - *SP* se representa el carácter o espacio en blanco
- **Caracteres geométricos o gráficos:**
  - Son símbolos o módulos con los que se pueden representar figuras (o iconos), bordes de cuadros, etc. Ejemplos:

    ♣ ♦ ♥ ♠ α β ⌠ ⌉ ⌡ ⌊ ⌈ ∑

6

## Caracteres de control

- **Representan órdenes de control:**
  - Carácter indicador de fin de texto (***ETX***)
  - Carácter de salto de página (***FF***)
  - carácter indicador de sincronización de una transmisión (***SYN***)
  - Carácter de fin de transmisión (***EOT***)
  - Emisión de un pitido en un terminal (***BEL***), etc.
- **Muchos de los caracteres de control son generados e insertados por la propia computadora.**

7

## Los códigos de textos se pueden definir de forma arbitraria

- **Supongamos *m=105* caracteres,**
  - como $2^6 < 105 < 2^7$ ; se necesitan *n=7 bits.*
  - De las 128 combinaciones posibles sólo se utilizarán 105(“*puntos de código*”). .
  - Arbitrariamente podemos asignar un carácter a cada código de 7 bits.
- **No obstante, existen códigos de E/S normalizados que son utilizados por diferentes constructores de computadores.**

8

## Algunos códigos normalizados

- SBCD (*6* bits).
- Fieldata (*6* bits)
- CÓDIGO EBCDIC (*8* bits)
- CÓDIGO ASCII (*7* y *8* bits)
- UNICODE (*16* a *24* bits)

9

## ASCII (ANSI-X3.4, 1968, 7 bits)

| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 00 | 0 | NUL | SOH | STX | ETX | EOT | ENQ | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 10 | 16 | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 20 | 32 | SP | ! | " | # | $ | % | & | ' | ( | ) | * | + | , | - | . | / |
| 30 | 48 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 40 | 64 | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 50 | 80 | P | Q | R | S | T | U | V | W | X | Y | Z | [ | \ | ] | ^ | _ |
| 60 | 96 | ` | a | b | c | d | e | f | g | h | i | j | k | L | m | n | o |
| 70 | 112 | p | q | r | s | t | u | V | w | x | y | z | { | \| | } | ~ | DEL |

**j→60 + A = 6A = 110 1010**

**Se suele añadir un bit de paridad**

10

## ASCII (ANSI-X3.4, 1968, 7 bits). Valor decimal del punto de código

| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 00 | 0 | NUL | SOH | STX | ETX | EOT | ENQ | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 10 | 16 | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 20 | 32 | SP | ! | " | # | $ | % | & | ' | ( | ) | * | + | , | - | . | / |
| 30 | 48 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 40 | 64 | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 50 | 80 | P | Q | R | S | T | U | V | W | X | Y | Z | [ | \ | ] | ^ | _ |
| 60 | 96 | ` | a | b | c | d | e | f | g | h | i | j | k | L | m | n | o |
| 70 | 112 | p | q | r | s | t | u | V | w | x | y | z | { | \| | } | ~ | DEL |

**Valor decimal: j→ 96 +10 =106**

11

## ASCII (ANSI-X3.4, 1968, 7 bits), otra forma de presentación

| Hex. | Dec. | | Hex. | Dec. | | Hex. | Dec. | | Hex. | Dec. | | Hex. | Dec. | | Hex. | Dec. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 032 | | 30 | 048 | 0 | 40 | 064 | @ | 50 | 080 | P | 60 | 096 | ` | 70 | 112 | p |
| 21 | 033 | ! | 31 | 049 | 1 | 41 | 065 | A | 51 | 081 | Q | 61 | 097 | a | 71 | 113 | q |
| 22 | 034 | " | 32 | 050 | 2 | 42 | 066 | B | 52 | 082 | R | 62 | 098 | b | 72 | 114 | r |
| 23 | 035 | # | 33 | 051 | 3 | 43 | 067 | C | 53 | 083 | S | 63 | 099 | c | 73 | 115 | s |
| 24 | 036 | $ | 34 | 052 | 4 | 44 | 068 | D | 54 | 084 | T | 64 | 100 | d | 74 | 116 | t |
| 25 | 037 | % | 35 | 053 | 5 | 45 | 069 | E | 55 | 085 | U | 65 | 101 | e | 75 | 117 | u |
| 26 | 038 | & | 36 | 054 | 6 | 46 | 070 | F | 56 | 086 | V | 66 | 102 | f | 76 | 118 | v |
| 27 | 039 | ' | 37 | 055 | 7 | 47 | 071 | G | 57 | 087 | W | 67 | 103 | g | 77 | 119 | w |
| 28 | 040 | ( | 38 | 056 | 8 | 48 | 072 | H | 58 | 088 | X | 68 | 104 | h | 78 | 120 | x |
| 29 | 041 | ) | 39 | 057 | 9 | 49 | 073 | I | 59 | 089 | Y | 69 | 105 | i | 79 | 121 | y |
| 2A | 042 | * | 3A | 058 | : | 4A | 074 | J | 5A | 090 | Z | 6A | 106 | j | 7A | 122 | z |
| 2B | 043 | + | 3B | 059 | ; | 4B | 075 | K | 5B | 091 | [ | 6B | 107 | k | 7B | 123 | { |
| 2C | 044 | , | 3C | 060 | < | 4C | 076 | L | 5C | 092 | \ | 6C | 108 | l | 7C | 124 | \| |
| 2D | 045 | - | 3D | 061 | = | 4D | 077 | M | 5D | 093 | ] | 6D | 109 | m | 7D | 125 | } |
| 2E | 046 | . | 3E | 062 | > | 4E | 078 | N | 5E | 094 | ^ | 6E | 110 | n | 7E | 126 | ~ |
| 2F | 047 | / | 3F | 063 | ? | 4F | 079 | O | 5F | 095 | _ | 6F | 111 | o | 7F | 127 | <d> |

12

## Caracteres de control. ASCII (ANSI-X3.4, 1968, 7 bits)

| | | | |
|---|---|---|---|
| NUL | Nulo | DC1 | Control de dispositivo 1 |
| SOH | Comienzo de cabecera | DC2 | Control de dispositivo 2 |
| STX | Comienzo de texto | DC3 | Control de dispositivo 3 |
| ETX | Final de texto | DC4 | Control de dispositivo 4 |
| EOT | Fin de transmisión | NAK | Acuse de recibo negativo |
| ENQ | Petición, consulta | SYN | Sincronización |
| ACK | Acuse de recibo | ETB | Final de bloque de transmisión |
| BEL | Pitido | CAN | Anulación |
| BS | Retroceso de 1 espacio | EM | Fin de soporte (cinta, etc.) |
| HT | Tabulación horizontal | SUB | Sustituir |
| LF | Saltar a línea siguiente | ESC | Escape |
| VT | Tabulación vertical | FS | Separador de fichero |
| FF | Alimentación de hoja | GS | Separador de grupo |
| CR | Retorno de carro | RS | Separador de registro |
| SO | Fuera de código | US | Separador de campo |
| SI | Dentro de código | DEL | Borrar, suprimir |
| DLE | Escape del enlace de datos | | |

13

## ASCII (Ampliaciones)

| Denominación | Estándar | Área geográfica |
|---|---|---|
| Latín-1 | ISO 8859-1 | Oeste y Europa del este |
| Latín-2 | ISO 8859-2 | Europa central y del este |
| Latín-3 | ISO 8859-3 | Europa sur, maltés y esperanto |
| Latín-4 | ISO 8859-4 | Europa norte |
| Alfabeto latín/cirílico | ISO 8859-5 | Lenguajes eslavos |
| Alfabeto latín/árabe | ISO 8859-6 | Lenguajes arábigos |
| Alfabeto latín/griego | ISO 8859-7 | Griego moderno |
| Alfabeto latín/hebraico | ISO 8859-8 | Hebreo y Yiddish |
| Latín-5 | ISO 8859-9 | Turco |
| Latín-6 | ISO 8859-10 | Nórdico (Sámi, Inuit e islandés) |
| Alfabeto Latín/Thai | ISO 8859-11 | Lenguaje Thai |
| Latín-7 | ISO 8859-13 | Báltico *Rim* |
| Latín-8 | ISO 8859-14 | Céltico |
| Latín-9 *(alias Latín-0)* | ISO 8859-15 | Latín 1 con ligeras modificaciones (símbolo €) |

14

## ASCII (ISO 8859-1, Latín 1)

**£ = A3**

| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 00 | 0 | NUL | SOH | STX | ETX | EOT | ENQ | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 10 | 16 | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 20 | 32 | SP | ! | " | # | $ | % | & | ' | ( | ) | * | + | , | - | . | / |
| 30 | 48 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 40 | 64 | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 50 | 80 | P | Q | R | S | T | U | V | W | X | Y | Z | [ | \ | ] | ^ | _ |
| 60 | 96 | ` | a | b | c | d | E | f | g | h | I | j | k | L | m | N | o |
| 70 | 112 | p | q | r | s | t | U | v | w | x | Y | z | { | \| | } | ~ | DEL |
| 80 | 128 | | | | | | | | | | | | | | | | |
| 90 | 144 | | | | | | | | | | | | | | | | |
| A0 | 160 | | ¡ | ¢ | £ | ¤ | ¥ | ¦ | § | ¨ | © | ª | « | ¬ | - | ® | ¯ |
| B0 | 176 | ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ |
| C0 | 192 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï |
| D0 | 208 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | ß |
| E0 | 224 | à | á | â | ã | ä | Å | Æ | ç | è | é | ê | ë | Ì | í | Î | ï |
| F0 | 240 | ð | ñ | ò | ó | ô | Õ | Ö | ÷ | ø | ù | ú | û | Ü | ý | Þ | ÿ |

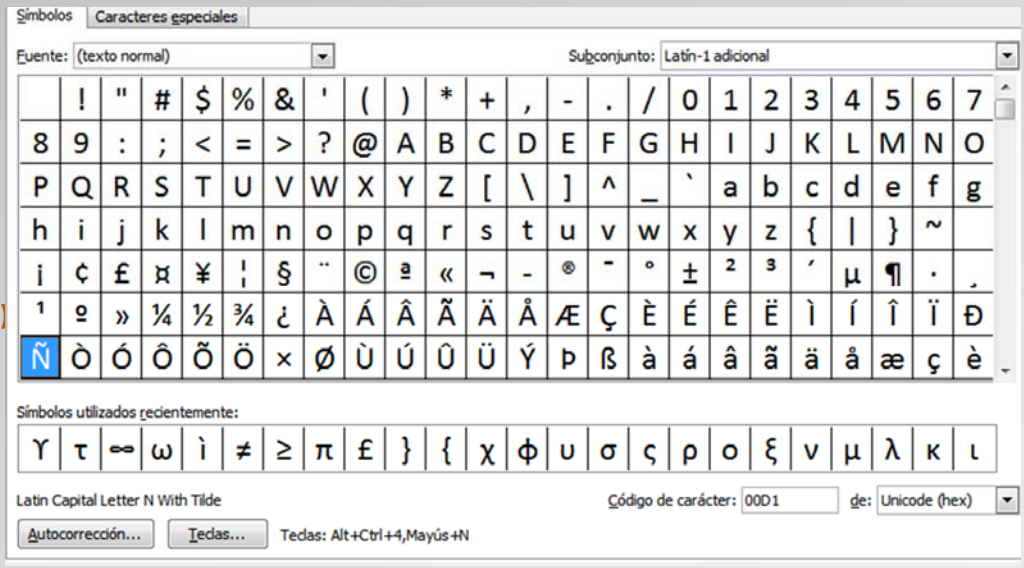
15

## Ejemplos de uso de la tecla Alt en combinación con valores decimales dados en las teclas numéricas

µ ~ Å @ ® ☐ ☐ ☐ ☐ ☐ ☐ € ☐ Ï µ ~ ☐

16

Insertar caracteres y símbolos que no están en el teclado con Microsoft Word (*insertar/símbolo*)




17

## Ejercicio: examinar el contenido binario de un fichero de texto

- Crear con NOTEPAD un fichero de texto
- Abrirlo con el programa **hexedit**

```
 0  48 6f 79 20 68 61 63 65 20 62 75 65 6e 20 74 69  Hoy hace buen ti
10  65 6d 70 6f 20 79 20 75 6e 61 20 74 65 6d 70 65  empo y una tempe
20  72 61 74 75 72 61 20 64 65 20 32 35 20 67 72 61  ratura de 25 gra
30  64 6f 73 2e 20 41 41 41 41 20 61 61 61 61 20 62  dos. AAAA aaaa b
40  62 62 62 20 30 30 30 30 20 31 31 31 31 20 2e 2e  bbb 0000 1111 ..
50  2e 2e 00                                         ..
```

18

## Inconvenientes de los códigos tradicionales (SBCD, EBCDIC, ASCII, etc.)

- **Los símbolos codificados son insuficientes para representar los caracteres especiales que requieren numerosas aplicaciones.**
- **Los símbolos y códigos añadidos en las versiones ampliadas a 8 bits no están normalizados.**
- **Están basados en los caracteres latinos, existiendo otras culturas que utilizan otros símbolos muy distintos.**
  - Los lenguajes escritos de diversas culturas orientales, como la china, japonesa y coreana se basan en la utilización de ideogramas o símbolos que representan palabras, frases o ideas completas, siendo, por tanto, inoperantes los códigos que sólo codifican letras individuales.

19

## Unicode (ISO/IEC 10646)

- **Propuesto en por un consorcio de empresas y entidades que trata de hacer posible escribir aplicaciones que sean capaces de procesar texto de muy diversas culturas. Se busca**
  - Universalidad,
    - trata de cubrir la mayoría de lenguajes escritos existentes en la actualidad: Inicialmente **16 bits ⇒ 65.356 símbolos** (ASCII ampliado: 256 caracteres)
  - Unicidad,
    - a cada carácter se le asigna exactamente un único código (idiogramas con imagen distinta, tienen igual código), y
  - Uniformidad,
    - ya que ***inicialmente*** todos los símbolos se representan con un número fijo de bits (***16***).

20

## Asignación de posiciones (*puntos de código*) en el Plano Básico Multilingüe (*BPM*)

| *Zona* | *Códigos* | | *Símbolos codificados* | *Nº de caractere* |
|---|---|---|---|---|
| **A** | 0000 | 0000<br>00FF | Latín-1 | 256 |
| | | | otros alfabetos | 7.936 |
| | | 2000 | Símbolos generales y caracteres fonéticos chinos, japoneses y coreanos | 8.192 |
| **I** | 4000 | | Ideogramas | 24.576 |
| **O** | A000 | | Pendiente de asignación | 16.384 |
| R | E000<br>FFFF | | Caracteres locales y propios de los usuarios.<br>Compatibilidad con otros códigos | 8.192 |

21

## Subjonjuntos Unicode estándarizados

| *Rango Unicode* | *Se corresponde con* |
|---|---|
| 0000 a 007F | Latín Básico (00 a 7F), definidos en la norma ASCII ANSI-X3.4. |
| 0080 a 00FF | Suplemento Latín-1 (ISO 8859-1) |
| 0100 a 017F | Ampliación A de Latín |
| 0180 a 024F | Ampliación B del Latín |
| 0250 a 02AF | Ampliación del Alfabeto Fonético Internacional (IPA) |
| 02BF a 02FF | Espaciado de letras modificadoras |
| 0300 a 036F | Combinación de marcas diacríticas (tilde, acento grave, etc.) |
| 0370 a 03FF | Griego |
| 0400 a 04FF | Cirílico |
| 0530 a 058F | Armenio |
| 0590 a 05FF | Hebreo |
| 0600 a 06FF | Árabe |
| 0700 a 074F | Sirio |
| etc. | etc. |

22

## Con el tiempo se han ido realizando ampliaciones, incluyendo nuevos "planos"

- En el BPM hay asignados sólo 24.576 puntos de código para ideogramas. El diccionario de la RAE contiene unas 88.000 palabras; pero una persona no suele utilizar más de unas 11.000.
- En la actualidad (Unicode 5.2 , 2009) hay asignados o reservados 17 planos → 17 x 216 = 1.114.112 puntos de código dentro del rango de 0000 a 10FFFF .
- En general, un punto Unicode se referencia escribiendo "U+" seguido por su nº HEX.

| | | |
|---|---|---|
| Plano 0 | Basic Multilingual Plane (BMP) | 0000–FFFF |
| Plano 1: | Supplementary Multilingual Plane (SMP): | 10000–1FFFF |
| Plano 2 | Supplementary Ideographic Plane (SIP): | 20000–2FFFF |
| Plano 3–13 | Sin asignar | 30000–DFFFF |
| Plano 14: | Supplementary Special-purpose Plane (SSP) | E0000–EFFFF |
| Planos 15–16 | Supplementary Private Use Area (S PUA A/B) | F0000–10FFFF |

23

## Recodificaciones: UTF (*Unicode Transformation Format*) y UCS (*Universal Character Set*)

- **UTF-8: La forma más usada en la actualidad.**
  - Los caracteres más probables se recodifican con menos bits.
  - Los 128 primeros se codifican con tan sólo 1 byte (son los caracteres ASCII)
  - A partir de U+007F se codifican con de dos a cuatro octetos (bytes).

| Bits del punto de código | 1er. punto de código | Último punto de código | Bytes en secuencia | Byte 1 | Byte 2 | Byte 3 | Byte 4 |
|---|---|---|---|---|---|---|---|
| **7** | U+0000 | U+007F | 1 | 0xxxxxxx | | | |
| **11** | U+0080 | U+07FF | 2 | 110xxxxx | 10xxxxxx | | |
| **16** | U+0800 | U+FFFF | 3 | 1110xxxx | 10xxxxxx | 10xxxxxx | |
| **21** | U+10000 | U+1FFFFF | 4 | 11110xxx | 10xxxxxx | 10xxxxxx | 10xxxxxx |

24

# Ejemplo de recodificación UTF-8

| Ejemplo | UNICODE | | | UTF-8 | | |
|---|---|---|---|---|---|---|
| | HEX | Nº bits | Bits del punto de código | Nº bytes | binario | HEX |
| E | 0045 | 7 | 100 0101 | 1 | 0100 0101 | 45 |
| ñ | 00F1 | 11 | 00011-110001 | 2 | 1100 0011 1011 0001 | C3 B1 |
| € | 20AC | 16 | 0010-000010-101100 | 3 | 1110 0010 1000 0010 1010 1100 | E2 82 AC |

| Bits del punto de código | 1er. punto de código | Último punto de código | Bytes en secuencia | Byte 1 | Byte 2 | Byte 3 | Byte 4 |
|---|---|---|---|---|---|---|---|
| **7** | U+0000 | U+007F | 1 | 0xxxxxxx | | | |
| **11** | U+0080 | U+07FF | 2 | 110xxxxx | 10xxxxxx | | |
| **16** | U+0800 | U+FFFF | 3 | 1110xxxx | 10xxxxxx | 10xxxxxx | |
| **21** | U+10000 | U+1FFFFF | 4 | 11110xxx | 10xxxxxx | 10xxxxxx | 10xxxxxx |

25

# Cuestiones prácticas: compresión con UTF-8

- **Supongamos que tenemos un archivo de texto UNICODE con:**
  - 18.325 caracteres ASCII básicos (entre U+0000 y U+007F)
  - 127 caracteres comprendidos entre U+0080 y U+07FF
  - 14 caracteres comprendidos entre U+0800 y U+FFFF
- **¿Qué factor de compresión se tiene si se recodifican en UTF-8?**

26

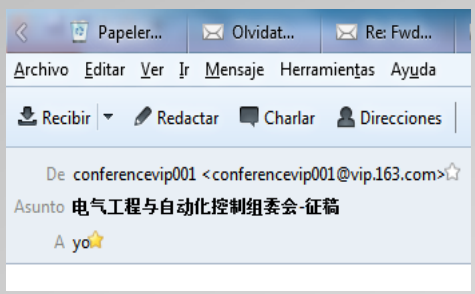
- **Capacidad en UNICODE:**
  - $C_{unicode}$ = (18.325 +127 + 14) · 2 Bytes = 18.466 ·2 = 36.932 Bytes
- **Capacidad en UTF-8**
  - $C_{UTF8}$ = (18.325·1 +127·2 + 14·3) Bytes = 18.621
- **Factor de compresión (definido en L2.1):**
  - $f_c = \frac{C_{unicode}}{C_{UTF8}} = \frac{36.932\ B}{18.621\ B} = 1{,}98$
  - Factor de compresión de 1,98 a 1

27

## Cuestiones prácticas: correo electrónico.

- **Gestor de correo electrónico**
  - Debe contener diferentes tablas de códigos, para poder seleccionar la correcta cuando llega un mensaje, y reconstruir y visualizar el original.
  - Debe detectar, de acuerdo con el protocolo de Internet, en la cabecera del mensaje el sistema de codificación con el que se ha enviado el mensaje.
  - **Torre de Babel** en Internet:
    - Mi gestor no contiene la tabla de códigos correspondiente al sistema en que se transmitió el mensaje original.
    - En la cabecera del mensaje recibido no se especifica el código en que se ha transmitido.
    - Mi gestor de correo electrónico no detecta adecuadamente la información del código en que se envió (y recibe) recibe el mensaje

28

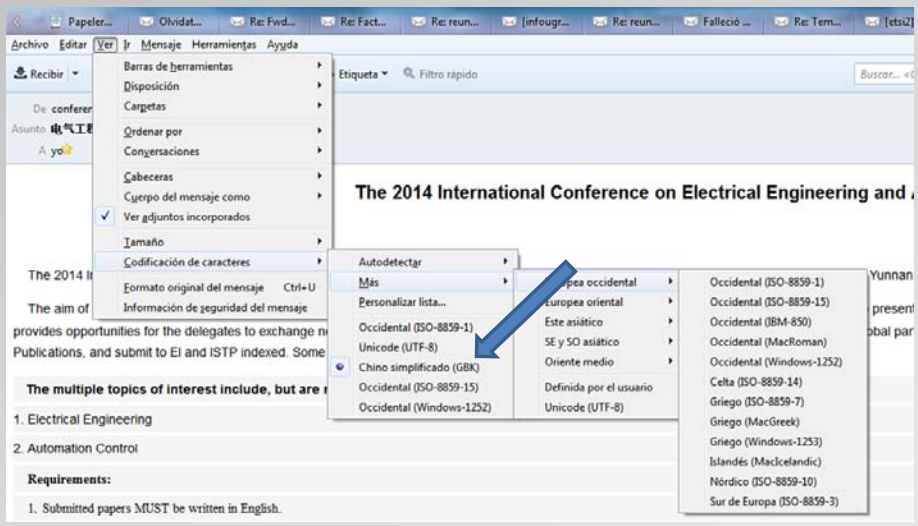




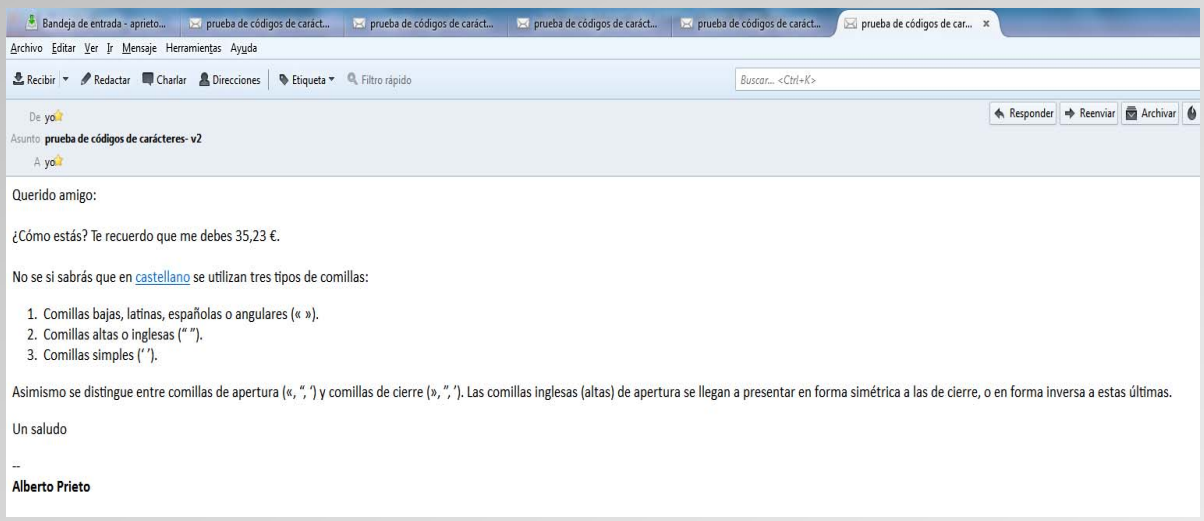
```
 -0.0 RCVD_IN_DNSWL_NONE     RBL: Sender listed at http://www.dnswl.org/, no
                             trust
                             [123.58.177.171 listed in list.dnswl.org]
  1.0 TVD_FROM_1             TVD_FROM_1
  0.0 FREEMAIL_FROM          Sender email is commonly abused enduser mail provider
                             (conferencevip001[at]vip.163.com)
 -0.0 SPF_PASS               SPF: sender matches SPF record
  0.2 FREEMAIL_ENVFROM_END_DIGIT Envelope-from freemail username ends in
                             digit (conferencevip001[at]vip.163.com)
  2.0 BAYES_80               BODY: Bayes spam probability is 80 to 95%
                             [score: 0.9274]
  0.0 HTML_MESSAGE           BODY: HTML included in message
  0.8 RDNS_NONE              Delivered to internal network by a host with no rDNS
  0.0 T_FILL_THIS_FORM_SHORT Fill in a short form with personal information
X-Spamina-History: valid
X-Spamina-Service-Type: pyme
X-Spamina-Destination: valid

------=_Part_214064_165620502.13969220
Content-Type: text/plain; charset=GBK
Content-Transfer-Encoding: base64

VGhlIDIwMTQgSW50ZXJuYXRpb25hbCBDb25mZXJlbmNlIG9uIEVsZWN0cmljYWwgRW5naW51ZXJp
bmcgYW5kIEF1dG9tYXRpb24gQ29udHJvbCBhRUVBQzIwMTRdCgogCgpUaGUgMjAxNCBJbnRlcm5h
dGlvbmFsIENvbmZlcmVuY2Ugb24gRWxlY3RyaWNhbCBFbmdpbmVlcmluZyBhbmQgQXV0b21hdGlv
biBDb250cm9sIFtFRUFDMjAxNF0gd2lsbCBiZSBoZWxkIGZyb20gQXVndXN0IDE2LTE3LDIwMTQg
aW4gS3VubWluZywgWXVubmFuLCBDaGluYS4KClRoZSBhaW0gb2YgRUVBQzIwMTQgaXMgdG8gcHJv
```



# Correo enviado desde China



# Mensaje enviado en UTF-8






# Mensaje anterior recibido e interpretado como si estuviese en ISO-8859-1. Se arregla configurando adecuadamente el gestor de correo (UTF-8)

## Resumen y conclusiones

- **L2.2 Representación digital de textos.**
  - Clasificación de los caracteres.
  - Códigos normalizados tradicionales: ASCII.
  - Unicode.
  - UTF-8.
  - Consideraciones prácticas.