

5. Tratamiento de formatos

| | |
|---|----|
| 5.1. Introducción..... | 2 |
| 5.2. Representación de la información textual: código de caracteres..... | 2 |
| 5.2.a Sistemas de numeración..... | 2 |
| 5.3. El lenguaje interno del ordenador y las unidades de medida en informática..... | 4 |
| 5.3.a. Los múltiplos del byte..... | 4 |
| 5.4. Representación de información no numérica..... | 4 |
| 5.4.a. Representación de texto..... | 5 |
| 5.5. Representación de la información textual: código de caracteres..... | 6 |
| 5.5.a. Conceptos básicos..... | 6 |
| 5.5.b. Algunas definiciones importantes..... | 7 |
| 5.5.c. Los códigos de caracteres más habituales..... | 7 |
| 5.5.d. Unicode..... | 15 |
| 5.5.e. Detección de la codificación de caracteres..... | 18 |
| 5.5.f. Cambio de la codificación de caracteres..... | 19 |
| 5.6. La representación de la información no textual..... | 20 |
| 5.6.a. Nombres de archivo y extensiones. Relación con el formato y la aplicación..... | 21 |
| 5.6.b. El formato HTML..... | 22 |
| 5.6.c. El XHTML..... | 25 |
| 5.6.d. Open Document..... | 26 |
| 5.6.e. Los formatos de documentos de Microsoft Word..... | 30 |
| Microsoft Word DOC (doc)..... | 30 |
| Microsoft Word 2003 XML (.xml)..... | 30 |
| Microsoft Word 2007/2010 XML (.docx)..... | 31 |
| 5.6.f. El formato LaTeX..... | 32 |
| 5.6.g. El formato DocBook..... | 33 |
| 5.6.h. El formato PDF..... | 34 |
| 5.7. XML..... | 36 |
| 5.7.a. Introducción..... | 36 |
| 5.7.b. Ejemplos sencillos de documentos XML..... | 38 |
| 5.7.c. Estructura de los documentos XML..... | 39 |
| 5.7.d. Los documentos XML bien formados..... | 40 |
| 5.7.e. Tecnologías asociadas: XSLT y XPath..... | 42 |
| 5.7.f. Editores de XML..... | 46 |
| 5.7.g. Traducción de documentos XML..... | 47 |
| 5.8. Los formatos XML utilizados en el mundo de la traducción..... | 49 |
| 5.8.a. Intercambio de memorias de traducción: TMX..... | 49 |
| 5.8.b. Intercambio de bases de datos terminológicas: TBX..... | 50 |
| 5.8.c. Intercambio de proyectos de traducción: XLIFF..... | 50 |
| 5.8.d. Intercambio de reglas de segmentación: SRX..... | 51 |
| 5.8.e. Métricas GILT: GMX..... | 52 |
| 6. Conclusiones..... | 52 |
| Para ampliar conocimientos..... | 53 |
| Tutoriales de W3Schools..... | 53 |
| Detección automática de lengua..... | 53 |
| Tablas de Unicode..... | 53 |
| Problemas de visualización de documentos relacionados con las fuentes..... | 55 |
| SC Unipad: Editor de Unicode..... | 56 |
| Bibliografía..... | 56 |

| | |
|---------------------------------|----|
| Anexo I. Entidades de html..... | 56 |
| Caràcters ASCII..... | 56 |
| Caràcters ISO-8859-1..... | 58 |
| Símbols ISO-8859-1..... | 60 |
| Símbols matemàtics..... | 61 |
| Lletres gregues..... | 61 |
| Altres símbols..... | 63 |

5.1. Introducción

El traductor menudo tiene que enfrentarse con problemas derivados de la gestión de los formatos y la codificación de caracteres de los ficheros que tiene que traducir. Por este motivo es imprescindible tener unas nociones sobre el funcionamiento básico de los formatos y las codificaciones. Con estos conocimientos evitaremos producir errores que a menudo hacen perder una cantidad de tiempo considerable.

El capítulo comienza con un repaso a una serie de conceptos muy básicos y probablemente conocidos para el lector, pero que son imprescindibles para comprender el resto del capítulo, especialmente en lo referente al tema de las codificaciones de caracteres.

En este capítulo también presentamos a fondo el formato XML, cada vez más utilizado, y por tanto, cada vez más traducido. Aprenderemos a traducir documentos XML con las herramientas de traducción más habituales. También dedicaremos un espacio a repasar los formatos basados en XML que se utilizan en el mundo de la traducción: TMX, TBX, SRX y XLIFF y que hemos ido viendo en los capítulos anteriores.

5.2. Representación de la información textual: código de caracteres

5.2.a Sistemas de numeración

Las personas estamos acostumbradas a utilizar un código de numeración decimal, es decir, un sistema que tiene un total de 10 símbolos: 0, 1, 2, 3, 4, 5, 6, 7, 8 y 9. Si queremos representar números mayores añadimos uno o más dígitos a la izquierda, por ejemplo, 10, 34, 234, 1234, etc.

Este sistema de numeración no es el único, existen otros, dos de los cuales son muy utilizados en informática. Uno de ellos es el *sistema binario*. El sistema binario sólo cuenta con dos símbolos, el 0 y el 1. Del mismo modo que en el sistema decimal, si necesitamos representar números mayores añadiremos uno o más dígitos a la izquierda, por ejemplo, 10, 11, 101, 11001011.

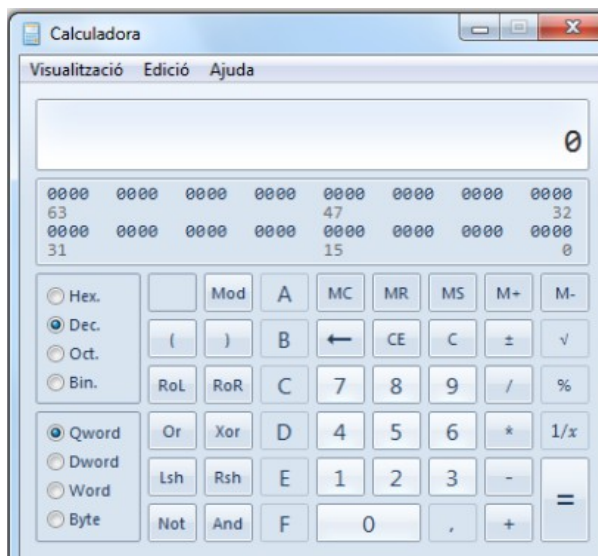
Otro sistema muy utilizado es el *sistema hexadecimal*, que cuenta con 16 símbolos: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E y F. Del mismo modo que con el sistema decimal y el sistema binario, si queremos representar números mayores añadiremos cifras a la izquierda: 10, 23, 2F, A10, BE3. En la tabla siguiente se listan algunos ejemplos de conversión entre los tres sistemas de numeración.

| Decimal | Binario | Hexadecimal |
|---------|---------|-------------|
| 0 | 0 | 0 |
| 1 | 01 | programazzz |
| 2 | 10 | 2 |
| 3 | 11 | 3 |
| 4 | 100 | 4 |
| 5 | 101 | 5 |
| 6 | 110 | 6 |
| 7 | 111 | 7 |
| 8 | 1000 | 8 |
| 9 | 1001 | 0 |
| 10 | 1010 | A |
| 11 | 1011 | B |
| 12 | 1100 | C |
| 13 | 1101 | D. |
| 14 | 1110 | E |
| 15 | 1111 | F |
| 16 | 10000 | 10 |

Otros ejemplos de conversiones: el número decimal 2003, es 11111010011 en binario y 7D3 en hexadecimal. El número hexadecimal F03A corresponde al decimal 61498 y al binario 1111000000111010.

Hay una serie de operaciones matemáticas no demasiado complicadas para pasar de un sistema de numeración a otro, pero no las estudiaremos en este capítulo. Por si algún día necesitas transformar de un sistema a otro (mira los ejercicios de este mismo capítulo), recuerda que hay muchas calculadoras científicas que realizan estas operaciones. Las calculadoras integradas en la mayoría de sistemas operativos son capaces de llevar a cabo estas conversiones. Abre la calculadora de Windows, la que aparece por defecto es una calculadora simple. Para cambiar de tipo de calculadora se puede ir a *Ver* y escoger la calculadora de *Programador* (si tu versión de Windows no dispone de este modo de calculadora, entonces tendrás que seleccionar la *Científica*):

Cuando se abre la calculadora normalmente está en modo Dec (decimal). Puedes introducir un número decimal y para pasarlo a binario, por ejemplo, sólo será necesario que seleccione el modo Bin (binario).



En este modo podrás escribir números en binario. También están disponibles los modos Hex (hexadecimal) y Oct (Octal, no lo hemos explicado, es un sistema de numeración que tiene 8 símbolos). Fíjate que cuando se ha seleccionado el modo Dec. las teclas numéricas del 0 al 9 están activadas, ya que todas estas son cifras válidas en este sistema de numeración. En cambio, cuando estamos en modo Bin. sólo quedan activadas las teclas 0 y 1, que son las únicas válidas. Cuando actives el modo Hex. además de las teclas numéricas del 0 al 9 se activan también las teclas de la A a la F, que en hexadecimal son cifras.

Ahora puedes hacer los siguiente ejercicio:

Pasar los siguientes números:

De decimal a hexadecimal: 23, 269, 62 165

De decimal a binario: 3, 15, 56, 258, 1645

De hexadecimal a decimal: 4, 1A, FE0, 10C0

5.3. El lenguaje interno del ordenador y las unidades de medida en informática

El ordenador internamente sólo trabaja con 0 y 1 (paso o no de corriente eléctrica), es decir trabaja con un sistema binario. En la información dada por un único dígito con sistema binario (0 o 1) la llamamos bit (el nombre proviene de *binary digit*). Para poder expresar una mayor cantidad de información los bits se agrupan en grupos de 8 que denominamos byte. Un byte puede tomar 256 valores diferentes (2⁸).

5.3.a. Los múltiplos del byte

Como la base de los cálculos en informática es el bit y éste sólo admite dos valores, todas las medidas se realizan con números que son potencias de 2. Un *kilobyte* (KB) son 1000 bytes (de hecho son 1024 ya que esta es la potencia de 2 más cercana al 1000). Un megabyte (MB) son 1.000.000 bytes (en realidad $1024 \times 1024 = 1.048.576$ bytes). Un gigabyte (GB) son 1.000.000.000 de bytes (en realidad $1024 \times 1024 \times 1024 = 1073741824$ bytes).

Para tener una idea de las capacidades de almacenamiento de diferentes unidades, ten en cuenta un disco duro estándar tiene actualmente entre 500GB y 1 TB en un CD_ROM caben hasta 700 MB, en un DVD 4,7 GB.

5.4. Representación de información no numérica

Como hemos visto los ordenadores trabajan con un código binario, que es capaz de representar números. Cuando trabajamos con ordenadores no únicamente queremos almacenar y trabajar con números, sino que también tenemos que ser capaces de procesar texto, sonido, imagen, etc. ¿Cómo podemos utilizar un código numérico para representar otro tipo de datos? A continuación presentamos la representación de texto.

5.4.a. Representación de texto

La idea básica para representar texto es asignar a cada carácter del conjunto que queremos representar un valor numérico. Si trabajamos con bytes de 8 bits podremos trabajar con 256 caracteres diferentes (8 bits, $2^8 = 256$). Aunque veremos los códigos de caracteres en detalle más adelante en este mismo capítulo, presentaremos ahora el código de caracteres ASCII (American Standard Code for Information Interchange) (7 bits, $2^7 = 128$) (fuente Wikipedia <http://en.wikipedia.org/wiki/Ascii>):

| Binary | Oct | Dec | Hex | Glyph | Binary | Oct | Dec | Hex | Glyph | Binary | Oct | Dec | Hex | Glyph |
|----------|-----|-----|-----|---------|----------|-----|-----|-----|-------|----------|-----|-----|-----|-------|
| 010 0000 | 040 | 32 | 20 | (space) | 100 0000 | 100 | 64 | 40 | @ | 110 0000 | 140 | 96 | 60 | ` |
| 010 0001 | 041 | 33 | 21 | ! | 100 0001 | 101 | 65 | 41 | A | 110 0001 | 141 | 97 | 61 | a |
| 010 0010 | 042 | 34 | 22 | " | 100 0010 | 102 | 66 | 42 | B | 110 0010 | 142 | 98 | 62 | b |
| 010 0011 | 043 | 35 | 23 | # | 100 0011 | 103 | 67 | 43 | C | 110 0011 | 143 | 99 | 63 | c |
| 010 0100 | 044 | 36 | 24 | \$ | 100 0100 | 104 | 68 | 44 | D | 110 0100 | 144 | 100 | 64 | d |
| 010 0101 | 045 | 37 | 25 | % | 100 0101 | 105 | 69 | 45 | E | 110 0101 | 145 | 101 | 65 | e |
| 010 0110 | 046 | 38 | 26 | & | 100 0110 | 106 | 70 | 46 | F | 110 0110 | 146 | 102 | 66 | f |
| 010 0111 | 047 | 39 | 27 | ' | 100 0111 | 107 | 71 | 47 | G | 110 0111 | 147 | 103 | 67 | g |
| 010 1000 | 050 | 40 | 28 | (| 100 1000 | 110 | 72 | 48 | H | 110 1000 | 150 | 104 | 68 | h |
| 010 1001 | 051 | 41 | 29 |) | 100 1001 | 111 | 73 | 49 | I | 110 1001 | 151 | 105 | 69 | i |
| 010 1010 | 052 | 42 | 2A | * | 100 1010 | 112 | 74 | 4A | J | 110 1010 | 152 | 106 | 6A | j |
| 010 1011 | 053 | 43 | 2B | + | 100 1011 | 113 | 75 | 4B | K | 110 1011 | 153 | 107 | 6B | k |
| 010 1100 | 054 | 44 | 2C | , | 100 1100 | 114 | 76 | 4C | L | 110 1100 | 154 | 108 | 6C | l |
| 010 1101 | 055 | 45 | 2D | - | 100 1101 | 115 | 77 | 4D | M | 110 1101 | 155 | 109 | 6D | m |
| 010 1110 | 056 | 46 | 2E | . | 100 1110 | 116 | 78 | 4E | N | 110 1110 | 156 | 110 | 6E | n |
| 010 1111 | 057 | 47 | 2F | / | 100 1111 | 117 | 79 | 4F | O | 110 1111 | 157 | 111 | 6F | o |
| 011 0000 | 060 | 48 | 30 | 0 | 101 0000 | 120 | 80 | 50 | P | 111 0000 | 160 | 112 | 70 | p |
| 011 0001 | 061 | 49 | 31 | 1 | 101 0001 | 121 | 81 | 51 | Q | 111 0001 | 161 | 113 | 71 | q |
| 011 0010 | 062 | 50 | 32 | 2 | 101 0010 | 122 | 82 | 52 | R | 111 0010 | 162 | 114 | 72 | r |
| 011 0011 | 063 | 51 | 33 | 3 | 101 0011 | 123 | 83 | 53 | S | 111 0011 | 163 | 115 | 73 | s |
| 011 0100 | 064 | 52 | 34 | 4 | 101 0100 | 124 | 84 | 54 | T | 111 0100 | 164 | 116 | 74 | t |
| 011 0101 | 065 | 53 | 35 | 5 | 101 0101 | 125 | 85 | 55 | U | 111 0101 | 165 | 117 | 75 | u |
| 011 0110 | 066 | 54 | 36 | 6 | 101 0110 | 126 | 86 | 56 | V | 111 0110 | 166 | 118 | 76 | v |
| 011 0111 | 067 | 55 | 37 | 7 | 101 0111 | 127 | 87 | 57 | W | 111 0111 | 167 | 119 | 77 | w |
| 011 1000 | 070 | 56 | 38 | 8 | 101 1000 | 130 | 88 | 58 | X | 111 1000 | 170 | 120 | 78 | x |
| 011 1001 | 071 | 57 | 39 | 9 | 101 1001 | 131 | 89 | 59 | Y | 111 1001 | 171 | 121 | 79 | y |
| 011 1010 | 072 | 58 | 3A | : | 101 1010 | 132 | 90 | 5A | Z | 111 1010 | 172 | 122 | 7A | z |
| 011 1011 | 073 | 59 | 3B | ; | 101 1011 | 133 | 91 | 5B | [| 111 1011 | 173 | 123 | 7B | { |
| 011 1100 | 074 | 60 | 3C | < | 101 1100 | 134 | 92 | 5C | \ | 111 1100 | 174 | 124 | 7C | |
| 011 1101 | 075 | 61 | 3D | = | 101 1101 | 135 | 93 | 5D |] | 111 1101 | 175 | 125 | 7D | } |
| 011 1110 | 076 | 62 | 3E | > | 101 1110 | 136 | 94 | 5E | ^ | 111 1110 | 176 | 126 | 7E | ~ |
| 011 1111 | 077 | 63 | 3F | ? | 101 1111 | 137 | 95 | 5F | _ | | | | | |

Podemos representar el código ASCII de una manera mucho más compacta. Por ejemplo, el carácter 'z' corresponde al código hexadecimal 7A, que en decimal equivale a 122.

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

Ejercicio:

Mirando esta tabla (no la de la página anterior, utilízala únicamente para comprobar la respuesta) escribe el código hexadecimal y decimal de los caracteres siguientes:

@:

q:

M:

~:

Por cierto, si alguna vez tienes dificultad para teclear algún carácter (ocurre a menudo con ~) en Windows puede hacer Alt + código decimal del carácter. Por ejemplo, para hacer la ~ puedes escribir 126 manteniendo pulsada la tecla Alt.

5.5. Representación de la información textual: código de caracteres

En la sección anterior hemos visto que podemos representar caracteres asignando un código numérico a cada carácter del conjunto que queremos representar. En esta unidad veremos a fondo todos los aspectos relacionados con los diferentes códigos de caracteres. Dedicaremos también una especial atención a Unicode. Aprenderemos también a determinar en qué código de caracteres está escrito un documento ya transformar códigos de caracteres.

5.5.a. Conceptos básicos

En informática como norma general los datos están representados como bytes. Un octeto es una unidad de información formada por 8 bits y que puede representar un valor numérico comprendido entre 0 y 255 ($2^8 = 256$). El concepto de octeto está muy relacionado con el concepto de byte.

Se pueden establecer diferentes convenciones sobre cómo un octeto o una secuencia de bytes representa un dato en concreto. Por ejemplo, bajo ciertos estándares, cuatro octetos consecutivos a menudo representan una unidad que representa un número real. En esta asignatura estamos interesados en la representación de caracteres. En el caso más sencillo, y que se utiliza muy a menudo, es que un octeto representa un carácter según una tabla de correspondencia. La interpretación correcta supone que se conoce el código de caracteres que se utiliza. Más adelante veremos qué técnicas existen para poder determinar la codificación de caracteres de un documento.

5.5.b. Algunas definiciones importantes

En este apartado intentaremos definir algunos conceptos importantes. La denominación que se utiliza no es universal ya menudo conduce a errores:

- **Repertorio de caracteres** (*character repertoire*): es el conjunto de caracteres diferentes a representar.
- **Código de caracteres** (*character code*): es una correspondencia, normalmente presentada en forma tabular, entre los caracteres de un repertorio de caracteres y un conjunto de números enteros positivos. Es decir, se asigna un código numérico único a cada carácter del repertorio.
- **Codificación de caracteres** (*character encoding*): Es un método (algoritmo) para representar los caracteres digitalmente haciendo una correspondencia entre las secuencias de códigos de caracteres y las secuencias de bytes.

En el caso más simple, cada carácter le corresponde un número entero entre 0 y 255 y éste se utiliza como octeto. Naturalmente, esta posibilidad sólo funciona para repertorios de caracteres de como máximo 256 carácter (cantidad que no es suficiente para todas las lenguas, pensemos, por ejemplo, en el chino).

5.5.c. Los códigos de caracteres más habituales

En este apartado describiremos los códigos de caracteres más utilizados. Dejaremos para el siguiente apartado todo lo que hace referencia al Unicode. Aquí mostraremos los siguientes códigos:

- ASCII
- La familia ISO 8859
- Códigos de caracteres de Windows
- Códigos de caracteres de DOS



- Códigos de caracteres de Machintosh
- La familia KOI de códigos de caracteres cirílicos

Hay más códigos de caracteres. No es necesario exponerlos todos, sino entender bien el mecanismo de funcionamiento. Aprenderemos a reconocer otros códigos de caracteres y transformarlos en próximos apartados.

ASCII

Ya hemos visto este código de caracteres en la sección anterior. Como vimos, el código de caracteres ANSI (*American Standard Code for Information Interchange*) es un código de 7 bits (128 posiciones). Ahora estamos hablando de octetos o bytes de 8 bits, por lo tanto nos sobra un bit (el primero). Este primer bit se puede utilizar como bit de paridad o bien para disponer de 128 posiciones adicionales (128-255).

La tabla de caracteres correspondientes al ASCII es la siguiente:

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

La familia ISO 8859

Esta familia de códigos de caracteres está formada por varias partes, y cada una cubre los caracteres necesarios para algunas lenguas. Son códigos de caracteres de 8 bits (por lo tanto pueden codificar 256 caracteres). La parte baja (los 7 primeros bits, es decir, los 128 primeros caracteres, de la posición 0 a la 127) de las tablas de la familia ISO 8859 es exactamente igual a la del ASCII. La parte alta se utiliza para codificar los caracteres no incluidos en el latín básico.

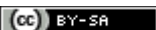
Veamos como ejemplo la tabla correspondiente a la ISO-8859-1 (extraída <http://czyborra.com/charsets/>), que representa los caracteres de las siguientes lenguas: afrikaans, albanés, euskera, bretón, catalán, corso, danés, inglés, feroés, gallego, alemán, islandés, indonesio, irlandés (nueva ortografía), italiano, latín (ortografía clásica básica), leonese, luxemburgués (ortografía clásica básica), malayo, gaélico manés, noruego (Bokmål and Nynorsk), occitano, portugués, retorrománico, gaélico escocés, español, swahili, sueco y valón.

| | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | - |
| B0 | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | BA | BB | BC | BD | BE | BF | ¿ |
| C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF | İ |
| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | DA | DB | DC | DD | DE | DF | B |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF | ı |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | FC | FD | FE | FF | ÿ |

Fijémonos en que la primera posición de esta tabla es la A0 (hexadecimal, en decimal es la 160, aunque esta primera posición está vacía). Hemos comentado que de la posición 0 a la 127 la table es exactamente igual que el ASCII. Por lo tanto de la posición 128 a la 159 tenemos un espacio vacío que no está ocupado por caracteres imprimibles.

En la siguiente tabla (adaptada de la Wikipedia http://en.wikipedia.org/wiki/ISO_8859 con las tablas de <http://czyborra.com/charsets/>) podemos observar las diferentes partes de la ISO 8859

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------------|--|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <p>Parte 1</p> | <p><i>Latin-1 Western European</i></p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>í</td><td>φ</td><td>£</td><td>¥</td><td>¥</td><td>ı</td><td>š</td><td>š</td><td>š</td><td>š</td><td>š</td><td>š</td><td>š</td><td>š</td><td>š</td></tr> <tr><td>E0</td><td>°</td><td>±</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>º</td><td>»</td><td>¼</td><td>½</td><td>¾</td></tr> <tr><td>C0</td><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ì</td><td>Í</td><td>Î</td></tr> <tr><td>D0</td><td>Ð</td><td>Ñ</td><td>Ò</td><td>Ó</td><td>Ô</td><td>Õ</td><td>Ö</td><td>×</td><td>Ø</td><td>Ù</td><td>Ú</td><td>Û</td><td>Ü</td><td>Ý</td><td>Þ</td></tr> <tr><td>E0</td><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td></tr> <tr><td>F0</td><td>ð</td><td>ñ</td><td>ò</td><td>ó</td><td>ô</td><td>õ</td><td>ö</td><td>÷</td><td>ø</td><td>ù</td><td>ú</td><td>û</td><td>ü</td><td>ý</td><td>þ</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | í | φ | £ | ¥ | ¥ | ı | š | š | š | š | š | š | š | š | š | E0 | ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | C0 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | D0 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | F0 | ð | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | í | φ | £ | ¥ | ¥ | ı | š | š | š | š | š | š | š | š | š | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ð | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 2</p> | <p><i>Latin-2 Central European</i></p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>À</td><td>Á</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td></tr> <tr><td>E0</td><td>°</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>º</td><td>»</td><td>¼</td><td>½</td><td>¾</td><td>¿</td></tr> <tr><td>C0</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td></tr> <tr><td>D0</td><td>Ð</td><td>Ñ</td><td>Ń</td><td>Ń</td><td>Ń</td><td>Ń</td><td>Ń</td><td>×</td><td>Ŕ</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ÿ</td><td>Ź</td></tr> <tr><td>E0</td><td>ř</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>č</td><td>č</td><td>ē</td><td>ē</td><td>ē</td><td>ē</td><td>ī</td><td>ī</td></tr> <tr><td>F0</td><td>đ</td><td>ñ</td><td>ń</td><td>ń</td><td>ń</td><td>ń</td><td>ń</td><td>÷</td><td>ř</td><td>ű</td><td>ű</td><td>ű</td><td>ű</td><td>ÿ</td><td>ź</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | À | Á | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | C0 | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | D0 | Ð | Ñ | Ń | Ń | Ń | Ń | Ń | × | Ŕ | Ū | Ū | Ū | Ū | Ÿ | Ź | E0 | ř | ā | ā | ā | ā | ā | ā | č | č | ē | ē | ē | ē | ī | ī | F0 | đ | ñ | ń | ń | ń | ń | ń | ÷ | ř | ű | ű | ű | ű | ÿ | ź |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | À | Á | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ń | Ń | Ń | Ń | Ń | × | Ŕ | Ū | Ū | Ū | Ū | Ÿ | Ź | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ř | ā | ā | ā | ā | ā | ā | č | č | ē | ē | ē | ē | ī | ī | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | đ | ñ | ń | ń | ń | ń | ń | ÷ | ř | ű | ű | ű | ű | ÿ | ź | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 3</p> | <p><i>Latin-3 South European</i></p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>Ħ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td><td>Ĳ</td></tr> <tr><td>E0</td><td>°</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>º</td><td>»</td><td>¼</td><td>½</td><td>¾</td><td>¿</td></tr> <tr><td>C0</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td></tr> <tr><td>D0</td><td>Ð</td><td>Ñ</td><td>Ń</td><td>Ń</td><td>Ń</td><td>Ń</td><td>Ń</td><td>×</td><td>Ŕ</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ÿ</td><td>Ź</td></tr> <tr><td>E0</td><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td></tr> <tr><td>F0</td><td>ñ</td><td>ò</td><td>ó</td><td>ô</td><td>õ</td><td>ö</td><td>÷</td><td>ø</td><td>ù</td><td>ú</td><td>û</td><td>ü</td><td>ý</td><td>ÿ</td><td>·</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | Ħ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | C0 | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | D0 | Ð | Ñ | Ń | Ń | Ń | Ń | Ń | × | Ŕ | Ū | Ū | Ū | Ū | Ÿ | Ź | E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | F0 | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | ÿ | · |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Ħ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | Ĳ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | Ā | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ń | Ń | Ń | Ń | Ń | × | Ŕ | Ū | Ū | Ū | Ū | Ÿ | Ź | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | ÿ | · | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 4</p> | <p><i>Latin-4 North European</i></p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>E0</td><td>°</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>º</td><td>»</td><td>¼</td><td>½</td><td>¾</td><td>¿</td></tr> <tr><td>C0</td><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ì</td><td>Í</td><td>Î</td></tr> <tr><td>D0</td><td>Ð</td><td>Ñ</td><td>Ò</td><td>Ó</td><td>Ô</td><td>Õ</td><td>Ö</td><td>×</td><td>Ø</td><td>Ù</td><td>Ú</td><td>Û</td><td>Ü</td><td>Ý</td><td>Þ</td></tr> <tr><td>E0</td><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td></tr> <tr><td>F0</td><td>ñ</td><td>ò</td><td>ó</td><td>ô</td><td>õ</td><td>ö</td><td>÷</td><td>ø</td><td>ù</td><td>ú</td><td>û</td><td>ü</td><td>ý</td><td>þ</td><td>·</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | C0 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | D0 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | F0 | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ | · |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ | · | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |



| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------------|-----------------------|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|---|---|---|---|---|---|---|---|---|---|---|----|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ë</td><td>Ì</td><td>Í</td></tr> <tr><td>E0</td><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ë</td><td>ì</td><td>í</td></tr> <tr><td>C0</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Ā</td><td>Æ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td><td>Ĉ</td></tr> <tr><td>D0</td><td>Ð</td><td>Ñ</td><td>Ō</td><td>Ō</td><td>Ō</td><td>Ō</td><td>×</td><td>Ø</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td></tr> <tr><td>E0</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>ā</td><td>æ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td><td>ĉ</td></tr> <tr><td>F0</td><td>đ</td><td>ñ</td><td>ō</td><td>ō</td><td>ō</td><td>ō</td><td>÷</td><td>ø</td><td>ū</td><td>ū</td><td>ū</td><td>ū</td><td>ū</td><td>ū</td><td>ū</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ë | Ì | Í | E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ë | ì | í | C0 | Ā | Ā | Ā | Ā | Ā | Ā | Æ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | D0 | Ð | Ñ | Ō | Ō | Ō | Ō | × | Ø | Ū | Ū | Ū | Ū | Ū | Ū | Ū | E0 | ā | ā | ā | ā | ā | ā | æ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | F0 | đ | ñ | ō | ō | ō | ō | ÷ | ø | ū | ū | ū | ū | ū | ū | ū |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ë | Ì | Í | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ë | ì | í | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Ā | Ā | Ā | Ā | Ā | Ā | Æ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | Ĉ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ō | Ō | Ō | Ō | × | Ø | Ū | Ū | Ū | Ū | Ū | Ū | Ū | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ā | ā | ā | ā | ā | ā | æ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | ĉ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | đ | ñ | ō | ō | ō | ō | ÷ | ø | ū | ū | ū | ū | ū | ū | ū | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 5</p> | <p>Latin/Cyrillic</p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>Ё</td><td>Ђ</td><td>Ѓ</td><td>Є</td><td>Ѕ</td><td>І</td><td>Ї</td><td>Ј</td><td>Љ</td><td>Њ</td><td>Ћ</td><td>Ќ</td><td>Ќ</td><td>Ў</td><td>Ѹ</td></tr> <tr><td>E0</td><td>А</td><td>Б</td><td>В</td><td>Г</td><td>Д</td><td>Е</td><td>Ж</td><td>З</td><td>И</td><td>Й</td><td>К</td><td>Л</td><td>М</td><td>Н</td><td>О</td></tr> <tr><td>C0</td><td>Р</td><td>С</td><td>Т</td><td>У</td><td>Ф</td><td>Х</td><td>Ц</td><td>Ч</td><td>Ш</td><td>Щ</td><td>Ъ</td><td>Ы</td><td>Ь</td><td>Э</td><td>Ю</td></tr> <tr><td>D0</td><td>а</td><td>б</td><td>в</td><td>г</td><td>д</td><td>е</td><td>ж</td><td>з</td><td>и</td><td>й</td><td>к</td><td>л</td><td>м</td><td>н</td><td>о</td></tr> <tr><td>E0</td><td>р</td><td>с</td><td>т</td><td>у</td><td>ф</td><td>х</td><td>ц</td><td>ч</td><td>ш</td><td>щ</td><td>ъ</td><td>ы</td><td>ь</td><td>э</td><td>ю</td></tr> <tr><td>F0</td><td>ё</td><td>ђ</td><td>ѓ</td><td>є</td><td>ѕ</td><td>і</td><td>ї</td><td>ј</td><td>љ</td><td>њ</td><td>ћ</td><td>ќ</td><td>ќ</td><td>ў</td><td>Ѹ</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | Ё | Ђ | Ѓ | Є | Ѕ | І | Ї | Ј | Љ | Њ | Ћ | Ќ | Ќ | Ў | Ѹ | E0 | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | C0 | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Ь | Э | Ю | D0 | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о | E0 | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | F0 | ё | ђ | ѓ | є | ѕ | і | ї | ј | љ | њ | ћ | ќ | ќ | ў | Ѹ |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Ё | Ђ | Ѓ | Є | Ѕ | І | Ї | Ј | Љ | Њ | Ћ | Ќ | Ќ | Ў | Ѹ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | А | Б | В | Г | Д | Е | Ж | З | И | Й | К | Л | М | Н | О | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Р | С | Т | У | Ф | Х | Ц | Ч | Ш | Щ | Ъ | Ы | Ь | Э | Ю | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | р | с | т | у | ф | х | ц | ч | ш | щ | ъ | ы | ь | э | ю | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ё | ђ | ѓ | є | ѕ | і | ї | ј | љ | њ | ћ | ќ | ќ | ў | Ѹ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 6</p> | <p>Latin/Arabic</p> | <table border="1"> <tr><td>A0</td><td></td><td></td><td>A4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>AC</td><td>AD</td><td></td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>BB</td><td></td><td></td><td>BF</td></tr> <tr><td></td><td>C1</td><td>C2</td><td>C3</td><td>C4</td><td>C5</td><td>C6</td><td>C7</td><td>C8</td><td>C9</td><td>CA</td><td>CB</td><td>CC</td><td>CD</td><td>CE</td><td>CF</td></tr> <tr><td>D0</td><td>ا</td><td>ب</td><td>ت</td><td>ث</td><td>ج</td><td>ح</td><td>خ</td><td>د</td><td>ذ</td><td>ر</td><td>ز</td><td>س</td><td>ش</td><td>ص</td><td>ض</td></tr> <tr><td>E0</td><td>ـ</td><td>ف</td><td>ق</td><td>ك</td><td>ل</td><td>م</td><td>ن</td><td>و</td><td>ي</td><td>ي</td><td>ي</td><td>ي</td><td>ي</td><td>ي</td><td>ي</td></tr> <tr><td>F0</td><td>ا</td><td>ب</td><td>ت</td><td>ث</td><td>ج</td><td>ح</td><td>خ</td><td>د</td><td>ذ</td><td>ر</td><td>ز</td><td>س</td><td>ش</td><td>ص</td><td>ض</td></tr> </table> | A0 | | | A4 | | | | | | | | | AC | AD | | | | | | | | | | | | | | | BB | | | BF | | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF | D0 | ا | ب | ت | ث | ج | ح | خ | د | ذ | ر | ز | س | ش | ص | ض | E0 | ـ | ف | ق | ك | ل | م | ن | و | ي | ي | ي | ي | ي | ي | ي | F0 | ا | ب | ت | ث | ج | ح | خ | د | ذ | ر | ز | س | ش | ص | ض | | | | | | | | | | | | | | | | |
| A0 | | | A4 | | | | | | | | | AC | AD | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | BB | | | BF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | ا | ب | ت | ث | ج | ح | خ | د | ذ | ر | ز | س | ش | ص | ض | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ـ | ف | ق | ك | ل | م | ن | و | ي | ي | ي | ي | ي | ي | ي | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ا | ب | ت | ث | ج | ح | خ | د | ذ | ر | ز | س | ش | ص | ض | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 7</p> | <p>Latin/Greek</p> | <table border="1"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td></td><td>ι</td><td>Ϸ</td><td>£</td><td></td><td></td><td>ι</td><td>ς</td><td></td><td>©</td><td></td><td>«</td><td>¬</td><td>-</td><td></td><td>-</td></tr> <tr><td>E0</td><td>±</td><td>²</td><td>³</td><td>⁴</td><td>⁵</td><td>⁶</td><td>⁷</td><td>⁸</td><td>⁹</td><td>Α</td><td>Β</td><td>Γ</td><td>Δ</td><td>Ε</td><td>Ζ</td></tr> <tr><td>C0</td><td>Α</td><td>Β</td><td>Γ</td><td>Δ</td><td>Ε</td><td>Ζ</td><td>Η</td><td>Θ</td><td>Ι</td><td>Κ</td><td>Λ</td><td>Μ</td><td>Ν</td><td>Ξ</td><td>Ο</td></tr> <tr><td>D0</td><td>Π</td><td>Ρ</td><td></td><td>Σ</td><td>Τ</td><td>Υ</td><td>Φ</td><td>Χ</td><td>Ψ</td><td>Ω</td><td>Ϊ</td><td>Ϋ</td><td>ά</td><td>έ</td><td>ή</td></tr> <tr><td>E0</td><td>ύ</td><td>α</td><td>β</td><td>γ</td><td>δ</td><td>ε</td><td>ζ</td><td>η</td><td>θ</td><td>ι</td><td>κ</td><td>λ</td><td>μ</td><td>ν</td><td>ξ</td></tr> <tr><td>F0</td><td>π</td><td>ρ</td><td>ς</td><td>σ</td><td>τ</td><td>υ</td><td>φ</td><td>χ</td><td>ψ</td><td>ω</td><td>ϊ</td><td>ϋ</td><td>ό</td><td>έ</td><td>ή</td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | ι | Ϸ | £ | | | ι | ς | | © | | « | ¬ | - | | - | E0 | ± | ² | ³ | ⁴ | ⁵ | ⁶ | ⁷ | ⁸ | ⁹ | Α | Β | Γ | Δ | Ε | Ζ | C0 | Α | Β | Γ | Δ | Ε | Ζ | Η | Θ | Ι | Κ | Λ | Μ | Ν | Ξ | Ο | D0 | Π | Ρ | | Σ | Τ | Υ | Φ | Χ | Ψ | Ω | Ϊ | Ϋ | ά | έ | ή | E0 | ύ | α | β | γ | δ | ε | ζ | η | θ | ι | κ | λ | μ | ν | ξ | F0 | π | ρ | ς | σ | τ | υ | φ | χ | ψ | ω | ϊ | ϋ | ό | έ | ή |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | ι | Ϸ | £ | | | ι | ς | | © | | « | ¬ | - | | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ± | ² | ³ | ⁴ | ⁵ | ⁶ | ⁷ | ⁸ | ⁹ | Α | Β | Γ | Δ | Ε | Ζ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Α | Β | Γ | Δ | Ε | Ζ | Η | Θ | Ι | Κ | Λ | Μ | Ν | Ξ | Ο | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | Π | Ρ | | Σ | Τ | Υ | Φ | Χ | Ψ | Ω | Ϊ | Ϋ | ά | έ | ή | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ύ | α | β | γ | δ | ε | ζ | η | θ | ι | κ | λ | μ | ν | ξ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | π | ρ | ς | σ | τ | υ | φ | χ | ψ | ω | ϊ | ϋ | ό | έ | ή | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | <p>Latin/Hebrew</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------------------|-----------------------------------|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|----|---|
| <p>Parte 8</p> | | <table border="1"> <tr><td>A0</td><td></td><td>A2</td><td>Φ</td><td>A3</td><td>£</td><td>A4</td><td>¥</td><td>A5</td><td>≠</td><td>A6</td><td>ı</td><td>A7</td><td>§</td><td>A8</td><td>..</td><td>A9</td><td>©</td><td>AR</td><td>×</td><td>AB</td><td>«</td><td>AC</td><td>¬</td><td>AD</td><td>-</td><td>AE</td><td>®</td><td>AF</td><td>-</td></tr> <tr><td>E0</td><td>°</td><td>B1</td><td>±</td><td>B2</td><td>²</td><td>B3</td><td>³</td><td>B4</td><td>ˆ</td><td>B5</td><td>μ</td><td>B6</td><td>¶</td><td>B7</td><td>•</td><td>B8</td><td>˙</td><td>B9</td><td>¹</td><td>BA</td><td>÷</td><td>BB</td><td>»</td><td>BC</td><td>¼</td><td>BD</td><td>½</td><td>BE</td><td>¾</td><td>BF</td><td></td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>DF</td><td>=</td></tr> <tr><td>E0</td><td>κ</td><td>E1</td><td>λ</td><td>E2</td><td>ι</td><td>E3</td><td>π</td><td>E4</td><td>ρ</td><td>E5</td><td>σ</td><td>E6</td><td>τ</td><td>E7</td><td>υ</td><td>E8</td><td>φ</td><td>E9</td><td>χ</td><td>EA</td><td>ψ</td><td>EB</td><td>ω</td><td>EC</td><td>δ</td><td>ED</td><td>ε</td><td>EE</td><td>ζ</td><td>EF</td><td>η</td></tr> <tr><td>F0</td><td>ι</td><td>F1</td><td>ο</td><td>F2</td><td>υ</td><td>F3</td><td>ϑ</td><td>F4</td><td>Ϙ</td><td>F5</td><td>ϙ</td><td>F6</td><td>Ϛ</td><td>F7</td><td>ϛ</td><td>F8</td><td>Ϝ</td><td>F9</td><td>ϝ</td><td>FA</td><td>Ϟ</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table> | A0 | | A2 | Φ | A3 | £ | A4 | ¥ | A5 | ≠ | A6 | ı | A7 | § | A8 | .. | A9 | © | AR | × | AB | « | AC | ¬ | AD | - | AE | ® | AF | - | E0 | ° | B1 | ± | B2 | ² | B3 | ³ | B4 | ˆ | B5 | μ | B6 | ¶ | B7 | • | B8 | ˙ | B9 | ¹ | BA | ÷ | BB | » | BC | ¼ | BD | ½ | BE | ¾ | BF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | DF | = | E0 | κ | E1 | λ | E2 | ι | E3 | π | E4 | ρ | E5 | σ | E6 | τ | E7 | υ | E8 | φ | E9 | χ | EA | ψ | EB | ω | EC | δ | ED | ε | EE | ζ | EF | η | F0 | ι | F1 | ο | F2 | υ | F3 | ϑ | F4 | Ϙ | F5 | ϙ | F6 | Ϛ | F7 | ϛ | F8 | Ϝ | F9 | ϝ | FA | Ϟ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| A0 | | A2 | Φ | A3 | £ | A4 | ¥ | A5 | ≠ | A6 | ı | A7 | § | A8 | .. | A9 | © | AR | × | AB | « | AC | ¬ | AD | - | AE | ® | AF | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | B1 | ± | B2 | ² | B3 | ³ | B4 | ˆ | B5 | μ | B6 | ¶ | B7 | • | B8 | ˙ | B9 | ¹ | BA | ÷ | BB | » | BC | ¼ | BD | ½ | BE | ¾ | BF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | DF | = | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | κ | E1 | λ | E2 | ι | E3 | π | E4 | ρ | E5 | σ | E6 | τ | E7 | υ | E8 | φ | E9 | χ | EA | ψ | EB | ω | EC | δ | ED | ε | EE | ζ | EF | η | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ι | F1 | ο | F2 | υ | F3 | ϑ | F4 | Ϙ | F5 | ϙ | F6 | Ϛ | F7 | ϛ | F8 | Ϝ | F9 | ϝ | FA | Ϟ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 9</p> | <p><i>Latin-5 Turkish</i></p> | <table border="1"> <tr><td>A0</td><td></td><td>A2</td><td>Φ</td><td>A3</td><td>£</td><td>A4</td><td>¥</td><td>A5</td><td>ı</td><td>A6</td><td>§</td><td>A8</td><td>..</td><td>A9</td><td>©</td><td>AR</td><td>×</td><td>AB</td><td>«</td><td>AC</td><td>¬</td><td>AD</td><td>-</td><td>AE</td><td>®</td><td>AF</td><td>-</td></tr> <tr><td>E0</td><td>°</td><td>B1</td><td>±</td><td>B2</td><td>²</td><td>B3</td><td>³</td><td>B4</td><td>ˆ</td><td>B5</td><td>μ</td><td>B6</td><td>¶</td><td>B7</td><td>•</td><td>B8</td><td>˙</td><td>B9</td><td>¹</td><td>BA</td><td>÷</td><td>BB</td><td>»</td><td>BC</td><td>¼</td><td>BD</td><td>½</td><td>BE</td><td>¾</td><td>BF</td><td>ç</td></tr> <tr><td>C0</td><td>ˆ</td><td>C1</td><td>ˆ</td><td>C2</td><td>ˆ</td><td>C3</td><td>ˆ</td><td>C4</td><td>ˆ</td><td>C5</td><td>ˆ</td><td>C6</td><td>æ</td><td>C7</td><td>ç</td><td>C8</td><td>ë</td><td>C9</td><td>ê</td><td>CA</td><td>ó</td><td>CB</td><td>»</td><td>CC</td><td>ı</td><td>CD</td><td>ı</td><td>CE</td><td>ı</td><td>CF</td><td>ı</td></tr> <tr><td>D0</td><td>ç</td><td>D1</td><td>ñ</td><td>D2</td><td>õ</td><td>D3</td><td>õ</td><td>D4</td><td>ô</td><td>D5</td><td>ö</td><td>D6</td><td>ö</td><td>D7</td><td>×</td><td>D8</td><td>ø</td><td>D9</td><td>ü</td><td>DA</td><td>ü</td><td>DB</td><td>û</td><td>DC</td><td>ü</td><td>DD</td><td>ı</td><td>DE</td><td>ş</td><td>DF</td><td>ß</td></tr> <tr><td>E0</td><td>â</td><td>E1</td><td>â</td><td>E2</td><td>â</td><td>E3</td><td>â</td><td>E4</td><td>ä</td><td>E5</td><td>â</td><td>E6</td><td>æ</td><td>E7</td><td>ç</td><td>E8</td><td>ë</td><td>E9</td><td>ê</td><td>EA</td><td>ó</td><td>EB</td><td>»</td><td>EC</td><td>ı</td><td>ED</td><td>ı</td><td>EE</td><td>ı</td><td>EF</td><td>ı</td></tr> <tr><td>F0</td><td>ç</td><td>F1</td><td>ñ</td><td>F2</td><td>õ</td><td>F3</td><td>õ</td><td>F4</td><td>ô</td><td>F5</td><td>ö</td><td>F6</td><td>ö</td><td>F7</td><td>÷</td><td>F8</td><td>ø</td><td>F9</td><td>ü</td><td>FA</td><td>ü</td><td>FB</td><td>û</td><td>FC</td><td>ü</td><td>FD</td><td>ı</td><td>FE</td><td>ş</td><td>FF</td><td>ÿ</td></tr> </table> | A0 | | A2 | Φ | A3 | £ | A4 | ¥ | A5 | ı | A6 | § | A8 | .. | A9 | © | AR | × | AB | « | AC | ¬ | AD | - | AE | ® | AF | - | E0 | ° | B1 | ± | B2 | ² | B3 | ³ | B4 | ˆ | B5 | μ | B6 | ¶ | B7 | • | B8 | ˙ | B9 | ¹ | BA | ÷ | BB | » | BC | ¼ | BD | ½ | BE | ¾ | BF | ç | C0 | ˆ | C1 | ˆ | C2 | ˆ | C3 | ˆ | C4 | ˆ | C5 | ˆ | C6 | æ | C7 | ç | C8 | ë | C9 | ê | CA | ó | CB | » | CC | ı | CD | ı | CE | ı | CF | ı | D0 | ç | D1 | ñ | D2 | õ | D3 | õ | D4 | ô | D5 | ö | D6 | ö | D7 | × | D8 | ø | D9 | ü | DA | ü | DB | û | DC | ü | DD | ı | DE | ş | DF | ß | E0 | â | E1 | â | E2 | â | E3 | â | E4 | ä | E5 | â | E6 | æ | E7 | ç | E8 | ë | E9 | ê | EA | ó | EB | » | EC | ı | ED | ı | EE | ı | EF | ı | F0 | ç | F1 | ñ | F2 | õ | F3 | õ | F4 | ô | F5 | ö | F6 | ö | F7 | ÷ | F8 | ø | F9 | ü | FA | ü | FB | û | FC | ü | FD | ı | FE | ş | FF | ÿ |
| A0 | | A2 | Φ | A3 | £ | A4 | ¥ | A5 | ı | A6 | § | A8 | .. | A9 | © | AR | × | AB | « | AC | ¬ | AD | - | AE | ® | AF | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | B1 | ± | B2 | ² | B3 | ³ | B4 | ˆ | B5 | μ | B6 | ¶ | B7 | • | B8 | ˙ | B9 | ¹ | BA | ÷ | BB | » | BC | ¼ | BD | ½ | BE | ¾ | BF | ç | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | ˆ | C1 | ˆ | C2 | ˆ | C3 | ˆ | C4 | ˆ | C5 | ˆ | C6 | æ | C7 | ç | C8 | ë | C9 | ê | CA | ó | CB | » | CC | ı | CD | ı | CE | ı | CF | ı | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | ç | D1 | ñ | D2 | õ | D3 | õ | D4 | ô | D5 | ö | D6 | ö | D7 | × | D8 | ø | D9 | ü | DA | ü | DB | û | DC | ü | DD | ı | DE | ş | DF | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | â | E1 | â | E2 | â | E3 | â | E4 | ä | E5 | â | E6 | æ | E7 | ç | E8 | ë | E9 | ê | EA | ó | EB | » | EC | ı | ED | ı | EE | ı | EF | ı | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ç | F1 | ñ | F2 | õ | F3 | õ | F4 | ô | F5 | ö | F6 | ö | F7 | ÷ | F8 | ø | F9 | ü | FA | ü | FB | û | FC | ü | FD | ı | FE | ş | FF | ÿ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 10</p> | <p><i>Latin-6 Nordic</i></p> | <table border="1"> <tr><td>A0</td><td></td><td>A2</td><td>Ē</td><td>A3</td><td>Ġ</td><td>A4</td><td>Ī</td><td>A5</td><td>ĩ</td><td>A6</td><td>Ķ</td><td>A7</td><td>š</td><td>A8</td><td>ł</td><td>A9</td><td>ð</td><td>AR</td><td>š</td><td>AB</td><td>ř</td><td>AC</td><td>ž</td><td>AD</td><td>-</td><td>AE</td><td>ū</td><td>AF</td><td>ŋ</td></tr> <tr><td>E0</td><td>°</td><td>B1</td><td>ā</td><td>B2</td><td>ē</td><td>B3</td><td>ġ</td><td>B4</td><td>ī</td><td>B5</td><td>ķ</td><td>B6</td><td>•</td><td>B7</td><td>ł</td><td>B8</td><td>đ</td><td>BA</td><td>š</td><td>BB</td><td>ř</td><td>BC</td><td>ž</td><td>BD</td><td>-</td><td>BE</td><td>ū</td><td>BF</td><td>ŋ</td></tr> <tr><td>C0</td><td>Ā</td><td>C1</td><td>Ā</td><td>C2</td><td>Ā</td><td>C3</td><td>Ā</td><td>C4</td><td>Ā</td><td>C5</td><td>Ā</td><td>C6</td><td>æ</td><td>C7</td><td>ı</td><td>C8</td><td>č</td><td>C9</td><td>ē</td><td>CA</td><td>ę</td><td>CB</td><td>ē</td><td>CC</td><td>ı</td><td>CD</td><td>ı</td><td>CE</td><td>ı</td><td>CF</td><td>ı</td></tr> <tr><td>D0</td><td>ð</td><td>D1</td><td>ō</td><td>D2</td><td>ō</td><td>D3</td><td>ō</td><td>D4</td><td>ô</td><td>D5</td><td>ö</td><td>D6</td><td>ö</td><td>D7</td><td>ü</td><td>D8</td><td>ø</td><td>D9</td><td>ų</td><td>DA</td><td>ų</td><td>DB</td><td>û</td><td>DC</td><td>ü</td><td>DD</td><td>ý</td><td>DE</td><td>þ</td><td>DF</td><td>ß</td></tr> <tr><td>E0</td><td>ā</td><td>E1</td><td>ā</td><td>E2</td><td>ā</td><td>E3</td><td>ā</td><td>E4</td><td>ä</td><td>E5</td><td>ā</td><td>E6</td><td>æ</td><td>E7</td><td>ı</td><td>E8</td><td>č</td><td>E9</td><td>ē</td><td>EA</td><td>ę</td><td>EB</td><td>ē</td><td>EC</td><td>ı</td><td>ED</td><td>ı</td><td>EE</td><td>ı</td><td>EF</td><td>ı</td></tr> <tr><td>F0</td><td>ð</td><td>F1</td><td>ō</td><td>F2</td><td>ō</td><td>F3</td><td>ō</td><td>F4</td><td>ô</td><td>F5</td><td>ö</td><td>F6</td><td>ö</td><td>F7</td><td>ü</td><td>F8</td><td>ø</td><td>F9</td><td>ų</td><td>FA</td><td>ų</td><td>FB</td><td>û</td><td>FC</td><td>ü</td><td>FD</td><td>ý</td><td>FE</td><td>þ</td><td>FF</td><td>κ</td></tr> </table> | A0 | | A2 | Ē | A3 | Ġ | A4 | Ī | A5 | ĩ | A6 | Ķ | A7 | š | A8 | ł | A9 | ð | AR | š | AB | ř | AC | ž | AD | - | AE | ū | AF | ŋ | E0 | ° | B1 | ā | B2 | ē | B3 | ġ | B4 | ī | B5 | ķ | B6 | • | B7 | ł | B8 | đ | BA | š | BB | ř | BC | ž | BD | - | BE | ū | BF | ŋ | C0 | Ā | C1 | Ā | C2 | Ā | C3 | Ā | C4 | Ā | C5 | Ā | C6 | æ | C7 | ı | C8 | č | C9 | ē | CA | ę | CB | ē | CC | ı | CD | ı | CE | ı | CF | ı | D0 | ð | D1 | ō | D2 | ō | D3 | ō | D4 | ô | D5 | ö | D6 | ö | D7 | ü | D8 | ø | D9 | ų | DA | ų | DB | û | DC | ü | DD | ý | DE | þ | DF | ß | E0 | ā | E1 | ā | E2 | ā | E3 | ā | E4 | ä | E5 | ā | E6 | æ | E7 | ı | E8 | č | E9 | ē | EA | ę | EB | ē | EC | ı | ED | ı | EE | ı | EF | ı | F0 | ð | F1 | ō | F2 | ō | F3 | ō | F4 | ô | F5 | ö | F6 | ö | F7 | ü | F8 | ø | F9 | ų | FA | ų | FB | û | FC | ü | FD | ý | FE | þ | FF | κ |
| A0 | | A2 | Ē | A3 | Ġ | A4 | Ī | A5 | ĩ | A6 | Ķ | A7 | š | A8 | ł | A9 | ð | AR | š | AB | ř | AC | ž | AD | - | AE | ū | AF | ŋ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ° | B1 | ā | B2 | ē | B3 | ġ | B4 | ī | B5 | ķ | B6 | • | B7 | ł | B8 | đ | BA | š | BB | ř | BC | ž | BD | - | BE | ū | BF | ŋ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | Ā | C1 | Ā | C2 | Ā | C3 | Ā | C4 | Ā | C5 | Ā | C6 | æ | C7 | ı | C8 | č | C9 | ē | CA | ę | CB | ē | CC | ı | CD | ı | CE | ı | CF | ı | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | ð | D1 | ō | D2 | ō | D3 | ō | D4 | ô | D5 | ö | D6 | ö | D7 | ü | D8 | ø | D9 | ų | DA | ų | DB | û | DC | ü | DD | ý | DE | þ | DF | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | ā | E1 | ā | E2 | ā | E3 | ā | E4 | ä | E5 | ā | E6 | æ | E7 | ı | E8 | č | E9 | ē | EA | ę | EB | ē | EC | ı | ED | ı | EE | ı | EF | ı | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | ð | F1 | ō | F2 | ō | F3 | ō | F4 | ô | F5 | ö | F6 | ö | F7 | ü | F8 | ø | F9 | ų | FA | ų | FB | û | FC | ü | FD | ý | FE | þ | FF | κ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 11</p> | <p>Latin/Thai</p> | <table border="1"> <tr><td></td><td>A1</td><td>A2</td><td>U</td><td>A3</td><td>U</td><td>A4</td><td>U</td><td>A5</td><td>U</td><td>A6</td><td>U</td><td>A7</td><td>U</td><td>A8</td><td>U</td><td>A9</td><td>U</td><td>AR</td><td>U</td><td>AB</td><td>U</td><td>AC</td><td>U</td><td>AD</td><td>U</td><td>AE</td><td>U</td><td>AF</td><td>U</td></tr> <tr><td>E0</td><td>U</td><td>E1</td><td>U</td><td>E2</td><td>U</td><td>E3</td><td>U</td><td>E4</td><td>U</td><td>E5</td><td>U</td><td>E6</td><td>U</td><td>E7</td><td>U</td><td>E8</td><td>U</td><td>EA</td><td>U</td><td>EB</td><td>U</td><td>EC</td><td>U</td><td>ED</td><td>U</td><td>EE</td><td>U</td><td>EF</td><td>U</td></tr> <tr><td>C0</td><td>C1</td><td>C2</td><td>C3</td><td>C4</td><td>C5</td><td>C6</td><td>C7</td><td>C8</td><td>C9</td><td>CA</td><td>CB</td><td>CC</td><td>CD</td><td>CE</td><td>CF</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>D0</td><td>D1</td><td>D2</td><td>D3</td><td>D4</td><td>D5</td><td>D6</td><td>D7</td><td>D8</td><td>D9</td><td>DA</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>E0</td><td>E1</td><td>E2</td><td>E3</td><td>E4</td><td>E5</td><td>E6</td><td>E7</td><td>E8</td><td>E9</td><td>EA</td><td>EB</td><td>EC</td><td>ED</td><td>EE</td><td>EF</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>F0</td><td>F1</td><td>F2</td><td>F3</td><td>F4</td><td>F5</td><td>F6</td><td>F7</td><td>F8</td><td>F9</td><td>FA</td><td>FB</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </table> | | A1 | A2 | U | A3 | U | A4 | U | A5 | U | A6 | U | A7 | U | A8 | U | A9 | U | AR | U | AB | U | AC | U | AD | U | AE | U | AF | U | E0 | U | E1 | U | E2 | U | E3 | U | E4 | U | E5 | U | E6 | U | E7 | U | E8 | U | EA | U | EB | U | EC | U | ED | U | EE | U | EF | U | C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF | | | | | | | | | | | | | | | | D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | DA | | | | | | | | | | | | | | | | | | | | | E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF | | | | | | | | | | | | | | | | F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | | | | | | | | | | | | | | | | | | | | | | | |
| | A1 | A2 | U | A3 | U | A4 | U | A5 | U | A6 | U | A7 | U | A8 | U | A9 | U | AR | U | AB | U | AC | U | AD | U | AE | U | AF | U | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | U | E1 | U | E2 | U | E3 | U | E4 | U | E5 | U | E6 | U | E7 | U | E8 | U | EA | U | EB | U | EC | U | ED | U | EE | U | EF | U | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | DA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | <p>Latin/Devanagar</p> | <p>The work in making a part of 8859 for Devanagari was</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------------------|---|--|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| <p>Parte 12</p> | <p>i</p> | <p>officially abandoned in 1997. ISCII and Unicode/ISO/IEC 10646 cover Devanagari.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 13</p> | <p><i>Latin-7 Baltic Rim</i></p> | <table border="1" data-bbox="708 436 1513 739"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td>°</td><td>±</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>²</td><td>»</td><td>¼</td><td>½</td><td>¾</td><td>æ</td></tr> <tr><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ì</td><td>Í</td><td>Î</td><td>Ï</td></tr> <tr><td>Š</td><td>Ń</td><td>Ņ</td><td>Ō</td><td>Õ</td><td>Ö</td><td>×</td><td>Ū</td><td>Ł</td><td>Ś</td><td>Ū</td><td>Ü</td><td>Ž</td><td>Ž</td><td>ß</td><td></td></tr> <tr><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td><td>ï</td></tr> <tr><td>š</td><td>ń</td><td>ņ</td><td>ō</td><td>õ</td><td>ö</td><td>÷</td><td>ū</td><td>ł</td><td>ś</td><td>ŭ</td><td>ü</td><td>ž</td><td>ž</td><td>'</td><td></td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | ² | » | ¼ | ½ | ¾ | æ | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | Š | Ń | Ņ | Ō | Õ | Ö | × | Ū | Ł | Ś | Ū | Ü | Ž | Ž | ß | | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | š | ń | ņ | ō | õ | ö | ÷ | ū | ł | ś | ŭ | ü | ž | ž | ' | |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | ² | » | ¼ | ½ | ¾ | æ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Š | Ń | Ņ | Ō | Õ | Ö | × | Ū | Ł | Ś | Ū | Ü | Ž | Ž | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| š | ń | ņ | ō | õ | ö | ÷ | ū | ł | ś | ŭ | ü | ž | ž | ' | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 14</p> | <p><i>Latin-8 Celtic</i></p> | <table border="1" data-bbox="708 889 1513 1191"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td>ƒ</td><td>ġ</td><td>Ġ</td><td>ġ</td><td>Ĥ</td><td>Ĥ</td><td>¶</td><td>Ɔ</td><td>ŵ</td><td>ŵ</td><td>ŵ</td><td>Ŷ</td><td>ŷ</td><td>Ŵ</td><td>Ŵ</td><td>Ŷ</td></tr> <tr><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ì</td><td>Í</td><td>Î</td><td>Ï</td></tr> <tr><td>Ŵ</td><td>Ņ</td><td>Ō</td><td>Ō</td><td>Ō</td><td>Ō</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ū</td><td>Ŷ</td><td>Ŷ</td><td>ß</td><td></td><td></td><td></td></tr> <tr><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td><td>ï</td></tr> <tr><td>ŵ</td><td>ņ</td><td>ō</td><td>ō</td><td>ō</td><td>ō</td><td>ū</td><td>ū</td><td>ū</td><td>ū</td><td>ŷ</td><td>ŷ</td><td>ß</td><td></td><td></td><td></td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | ƒ | ġ | Ġ | ġ | Ĥ | Ĥ | ¶ | Ɔ | ŵ | ŵ | ŵ | Ŷ | ŷ | Ŵ | Ŵ | Ŷ | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | Ŵ | Ņ | Ō | Ō | Ō | Ō | Ū | Ū | Ū | Ū | Ŷ | Ŷ | ß | | | | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | ŵ | ņ | ō | ō | ō | ō | ū | ū | ū | ū | ŷ | ŷ | ß | | | |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ƒ | ġ | Ġ | ġ | Ĥ | Ĥ | ¶ | Ɔ | ŵ | ŵ | ŵ | Ŷ | ŷ | Ŵ | Ŵ | Ŷ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Ŵ | Ņ | Ō | Ō | Ō | Ō | Ū | Ū | Ū | Ū | Ŷ | Ŷ | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ŵ | ņ | ō | ō | ō | ō | ū | ū | ū | ū | ŷ | ŷ | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 15</p> | <p>Latin-9</p> | <table border="1" data-bbox="708 1303 1513 1606"> <tr><td>A0</td><td>A1</td><td>A2</td><td>A3</td><td>A4</td><td>A5</td><td>A6</td><td>A7</td><td>A8</td><td>A9</td><td>AA</td><td>AB</td><td>AC</td><td>AD</td><td>AE</td><td>AF</td></tr> <tr><td>°</td><td>±</td><td>²</td><td>³</td><td>´</td><td>µ</td><td>¶</td><td>·</td><td>¸</td><td>¹</td><td>²</td><td>»</td><td>¼</td><td>½</td><td>¾</td><td>ß</td></tr> <tr><td>À</td><td>Á</td><td>Â</td><td>Ã</td><td>Ä</td><td>Å</td><td>Æ</td><td>Ç</td><td>È</td><td>É</td><td>Ê</td><td>Ë</td><td>Ì</td><td>Í</td><td>Î</td><td>Ï</td></tr> <tr><td>Ð</td><td>Ń</td><td>Ō</td><td>Ō</td><td>Ō</td><td>Ō</td><td>×</td><td>Ū</td><td>Ł</td><td>Ś</td><td>Ū</td><td>Ü</td><td>Ÿ</td><td>ß</td><td></td><td></td></tr> <tr><td>à</td><td>á</td><td>â</td><td>ã</td><td>ä</td><td>å</td><td>æ</td><td>ç</td><td>è</td><td>é</td><td>ê</td><td>ë</td><td>ì</td><td>í</td><td>î</td><td>ï</td></tr> <tr><td>ð</td><td>ń</td><td>ō</td><td>ō</td><td>ō</td><td>ō</td><td>÷</td><td>ū</td><td>ł</td><td>ś</td><td>ŭ</td><td>ü</td><td>ÿ</td><td>ß</td><td></td><td></td></tr> </table> | A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | ² | » | ¼ | ½ | ¾ | ß | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | Ð | Ń | Ō | Ō | Ō | Ō | × | Ū | Ł | Ś | Ū | Ü | Ÿ | ß | | | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | ð | ń | ō | ō | ō | ō | ÷ | ū | ł | ś | ŭ | ü | ÿ | ß | | |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ° | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | ² | » | ¼ | ½ | ¾ | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Ð | Ń | Ō | Ō | Ō | Ō | × | Ū | Ł | Ś | Ū | Ü | Ÿ | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ð | ń | ō | ō | ō | ō | ÷ | ū | ł | ś | ŭ | ü | ÿ | ß | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Parte 16</p> | <p><i>Latin-10 South-Eastern European</i></p> | <p>Intended for Albanian, Croatian, Hungarian, Italian, Polish, Romanian and Slovene, but also Finnish, French, German and Irish Gaelic (new orthography). The focus lies more on letters than symbols. The currency sign is replaced with the euro sign.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |



Códigos de caracteres de Windows

En ISO 8859-1 las posiciones 128-159 están reservadas para caracteres de control, es decir, corresponden a combinaciones que no representan caracteres gráficos. Los códigos de caracteres de Windows (WinLatin1 o Windows code page 1251) utiliza algunas de estas posiciones para caracteres imprimibles. Veamos aquí la tabla correspondiente (extraída de <http://czyborra.com/charsets/codepages.html>):

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| 80 | € | | 82 | , | 83 | f | 84 | „ | 85 | … | 86 | † | 87 | ‡ | 88 | ˆ | 89 | % | 9A | ˘ | 9B | < | 9C | œ | 9E | ˘ | 9F | ˘ | | | |
| | 91 | ı | 92 | ı | 93 | ıı | 94 | ıı | 95 | • | 96 | - | 97 | - | 98 | ~ | 99 | ™ | 9A | ˘ | 9B | > | 9C | œ | 9E | ˘ | 9F | ˘ | | | |
| A0 | A1 | ı | A2 | φ | A3 | £ | A4 | ¤ | A5 | ¥ | A6 | ı | A7 | š | A8 | ˙ | A9 | © | AA | ® | AB | « | AC | ¬ | AD | - | AE | ® | AF | - | |
| B0 | ø | B1 | ± | B2 | ² | B3 | ³ | B4 | - | B5 | μ | B6 | ¶ | B7 | · | B8 | ˘ | B9 | ı | BA | º | BB | » | BC | ¼ | BD | ½ | BE | ¾ | BF | ¿ |
| C0 | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï | | | | | | | | | | | | | | | |
| D0 | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | ß | | | | | | | | | | | | | | | |
| E0 | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï | | | | | | | | | | | | | | | |
| F0 | ä | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ | ÿ | | | | | | | | | | | | | | | |

Compara las posiciones en esta tabla y en la correspondiente de la ISO-8859-1 de algunos caracteres, por ejemplo: A, à, ç, ñ, ö. ¿Son las mismas?

Hay códigos de caracteres de Windows para otros alfabetos, concretamente:

- 1250 (Central Europe)
- 1251 (Cyrillic)
- 1252 (Latin I)
- 1253 (Greek)
- 1254 (Turkish)
- 1255 (Hebrew)
- 1256 (Arabic)
- 1257 (Baltic)
- 1258 (Vietnam)
- 874 (Thai)

La familia KOI de códigos de caracteres cirílicos

Ponemos como ejemplo esta familia para destacar que desde los inicios de la informática existían códigos de caracteres para otros alfabetos, como por ejemplo los caracteres cirílicos. KOI significa Код Обмена Информацией ("Kod Obmena Informatsiy" - Código para el Intercambio de Información). El primer KOI, llamado KOI-7 era un código de carácter de 7 bits que tenía sólo letras en mayúsculas:

KOI-7

| | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 2A | 2B | 2C | 2D | 2E | 2F |
| | ! | " | # | ¤ | % | & | ' | (|) | * | + | , | - | . | / |
| 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 3A | 3B | 3C | 3D | 3E | 3F |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 4A | 4B | 4C | 4D | 4E | 4F |
| @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 5A | 5B | 5C | 5D | 5E | 5F |
| P | Q | R | S | T | U | V | W | X | Y | Z | [| \ |] | ^ | _ |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 6A | 6B | 6C | 6D | 6E | 6F |
| Ю | А | Б | Ц | Д | Е | Ф | Г | Х | И | Й | К | Л | М | Н | О |
| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 7A | 7B | 7C | 7D | 7E | 7F |
| П | Я | Р | С | Т | У | Ж | В | Ь | Ы | Э | Ш | Щ | Ч | | |

El KOI-8

Esta es una versión de 8 bits, e incluye tanto caracteres en mayúsculas como en minúsculas. En la siguiente figura podemos ver la porción superior:

| | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF |
| Ю | а | б | ц | д | е | ф | г | х | и | й | к | л | м | н | о |
| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | DA | DB | DC | DD | DE | DF | |
| П | я | р | с | т | у | ж | в | ь | ы | э | ш | э | щ | ч | ъ |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF |
| Ю | А | Б | Ц | Д | Е | Ф | Г | Х | И | Й | К | Л | М | Н | О |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | FC | FD | FE | FF |
| П | Я | Р | С | Т | У | Ж | В | Ь | Ы | Э | Ш | Э | Щ | Ч | |

KOI-8 con ë (KOI8-R)

| | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | | | A3 | | | | | | | | | | | | |
| | | | B3 | | | | | | | | | | | | |
| | | | ë | | | | | | | | | | | | |
| | | | Ë | | | | | | | | | | | | |
| C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF |
| Ю | а | б | ц | д | е | ф | г | х | и | й | к | л | м | н | о |
| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | DA | DB | DC | DD | DE | DF | |
| П | я | р | с | т | у | ж | в | ь | ы | э | ш | э | щ | ч | ъ |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF |
| Ю | А | Б | Ц | Д | Е | Ф | Г | Х | И | Й | К | Л | М | Н | О |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | FC | FD | FE | FF |
| П | Я | Р | С | Т | У | Ж | В | Ь | Ы | Э | Ш | Э | Щ | Ч | Ъ |

5.5.d. Unicode

Introducción

En el apartado anterior hemos presentado una serie de códigos de caracteres que utilizan 8 bits. Esto da la posibilidad de codificar hasta 256 caracteres. Para muchos idiomas esto es suficiente, pero no para todos (piensa, por ejemplo en los caracteres chinos). Aunque pueda ser suficiente para muchos idiomas, hace que sea imposible guardar en un único archivo de texto (¡Atención! de texto, utilizando otros formatos sí que es posible) documentos multilingües para ciertas combinaciones de idiomas (por ejemplo, combinar en un único documento catalán y ruso). También se debe tener en cuenta que de vez en cuando aparecen nuevos símbolos (pensemos por ejemplo en el € del euro) que se deben ir incorporando al código de caracteres.

En el apartado anterior hemos visto unos cuantos códigos de caracteres de los muchos existentes. Esta gran cantidad de códigos de caracteres implica la dificultad de abrir un documento correctamente (aspecto que trataremos más adelante en este mismo capítulo) ya que la detección del código de caracteres no es totalmente automática.

Por estos motivos se intenta adoptar un código de caracteres universal. Este código de caracteres es el Unicode. Unicode utiliza más de 8 bits, por lo que puede codificar muchos más caracteres.

Originariamente se pensaba utilizar simplemente una codificación de 16 bits que proporciona la posibilidad de codificar más de 65.000 caracteres ($2^{16} = 65.536$). Aunque esta cifra es suficiente para codificar la mayoría de los miles de caracteres que se utilizan en las diferentes lenguas del mundo, el estándar Unicode ISO / IEC 10646 permite tres formas de codificación que utilizan un repertorio de caracteres común pero que permiten codificar alrededor de un millón más de caracteres. Esta cifra es suficiente para cubrir todas las necesidades de codificación conocidas, incluyendo todas las escrituras históricas del mundo y otros sistemas de notación.

Codificaciones de caracteres con Unicode

Existen diferentes maneras de codificar los caracteres con Unicode. La mayoría de ordenadores utilizan unidades mínimas de 8 bits. Si utilizamos más de 8 bits deberemos organizar la codificación de forma que utilicemos múltiples de 8 bits, es decir, más de un byte. El estándar Unicode define tres tipos de codificaciones que permite representar la información en un byte, dos bytes o cuatro bytes. Las tres codificaciones codifican el mismo repertorio de caracteres común y se puede pasar de una codificación a otra sin pérdida de datos:

- **UTF-8:** la codificación en bytes es de una longitud variable, desde 1 byte para los caracteres coincidentes con el ASCII.
- **UTF-16:** la codificación también es variable, pero o bien en dos bytes o bien en cuatro.
- **UTF-32:** todos los caracteres se codifican con cuatro bytes.

UTF-8

En el siguiente esquema podemos observar cómo se codifican los caracteres en UTF-8

| Bits | Última posición (HEX) | Número de caracteres | Byte 1 | Byte 2 | Byte 3 | Byte 4 |
|------|-----------------------|----------------------|----------|----------|----------|----------|
| 7 | U + 007F | 127 | 0xxxxxxx | | | |
| 11 | U + 07FF | 2.048 | 110xxxxx | 10xxxxxx | | |
| 16 | U + FFFF | 65.536 | 1110xxxx | 10xxxxxx | 10xxxxxx | |
| 21 | U + 1FFFFF | 2.097.152 | 11110xxx | 10xxxxxx | 10xxxxxx | 10xxxxxx |

Fijémonos en primer lugar, que no todos los bits de los diferentes bytes se utilizan para contener información, sino que algunos de ellos se utilizan para indicar cómo se deben leer estos bytes. Así, si el byte empieza por 0, indica que sólo se tiene que leer un byte (y por tanto quedan 7 bits para contener información y el número de caracteres posibles es de 127). La parte más baja del Unicode coincide plenamente con el ASCII. De esta manera se asegura la retrocompatibilidad con el ASCII. Esto también significa que un archivo que contenga sólo caracteres latinos básicos y cifras y puntuaciones (por ejemplo un texto en inglés) y que esté en codificación ASCII es exactamente igual que este mismo archivo en Unicode UTF-8.

Si se tienen que leer dos bytes, el primero empieza por 110 (y por tanto tiene disponibles 5 bits para información) y el segundo empieza por 10 (y tiene disponibles 6). Ten en cuenta que todos los bytes que no son el primero tienen la estructura 10xxxxxx. Generalizando, si un byte comienza por 1, quiere decir que tendrá que leer más de un byte y el número de bytes que hay que leer (incluyendo este primero) vendrá dado por el número de 1 del primero.

Fíjate también que en UTF-8, si disponemos de 4 bytes tenemos disponibles 21 bits para información y no 32 (8x4).

UTF-16

El UTF-16 tiene una longitud variable de 1 o 2 palabras de 16 bits, es decir, de 2 o 4 bytes. Está optimizado para representar los caracteres del *plano básico multilingüe* (BMP o *Basic Multilingual Plane*), es decir, los caracteres que se usan más frecuentemente en todas las lenguas (incluidos los caracteres chinos, japoneses y coreanos). Estos caracteres están inclusive en el rango U + 0000 a U + FFFF (0 a 65535) y cuando se limita a representar caracteres de este plano utiliza una longitud fija de 16 bits.

El UTF-16 produce una secuencia de unidades de 16 bits. Como estas unidades tienen 2 bytes de 8 bits, el orden de estos bits puede depender de la *endianess* (orden de los bytes) de la arquitectura del ordenador. Fijémonos en la siguiente tabla (extraída de la Wikipedia <http://en.wikipedia.org/wiki/Utf-16>) para explicar este concepto.

| Posición | Glifo [1] | Carácter | Unidades del código UTF-16 (hex) | Unidades del código UTF-16BE (hex) | Unidades del código UTF-16LE (hex) |
|----------|-----------|------------------------------------|----------------------------------|------------------------------------|------------------------------------|
| U+007A | z | LATIN SMALL LETTER Z | 007A | 00, 7A | 7A, 00 |
| U + 6C34 | ☐ | CJK UNIFIED IDEOGRAPH-6C34 (water) | 6C34 | 6C, 34 | 34, 6C |

La z minúscula, que está en la posición 007A, si se representa como *Big Endian* queda igual (007A), pero si se representa como *Little Endian* queda como 7A00. Si el sistema falla en determinar la *endianess* podrá confundir la z con el símbolo ☐.

Para ayudar a reconocer el orden de los bytes (*el endianess*) UTF-16 permite una Marca de Orden de Byte (BOM - *Byte Order Mark*), que un valor de U + FEFF y que precede el valor codificado real (U + FEFF es el espacio invisible de ancho cero / carácter ZWNBSP - *invisible zero-width non-breaking space*). Si la *endianess* del decodificador coincide con la del codificador, el decodificador detecta el valor correcto de 0xFEFF; pero si no es así se detecta el valor U + FFFE que está reservado para esta función. Este error permite corregir *el endianess* para el resto de valores. Si no hay BOM la norma RFC 2781 dice que se ha de suponer una codificación *big-endian*, pero como Windows utiliza *little-endian* por defecto, muchas aplicaciones suponen esta opción por defecto. En ausencia de BOM, para determinar la *endianess* suele buscar el carácter espacio (U + 0020) que es muy frecuente en los textos de la mayoría de lenguas.

UTF-32

El UTF-32 utiliza 32 bits (4 bytes) para todos los caracteres. En UTF-32 los caracteres se representan directamente según su posición Unicode. La ventaja principal es que el algoritmo para leer los archivos es muy simple y rápido, ya que simplemente tiene que ir leyendo 4 bytes en 4 bytes. El principal inconveniente es que es ineficiente en cuanto a espacio, ya que utiliza el doble que el UTF-16 y hasta 4 veces más que el UTF-8 (aunque esto dependerá del carácter a representar).

En la siguiente tabla (extraída de http://en.wikipedia.org/wiki/Comparison_of_Unicode_encodings) podemos observar el número de bytes empleado para cada codificación según los rangos de los códigos de caracteres.

| Rango de códigos (hexadecimal) | UTF-8 | UTF-16 | UTF-32 |
|--------------------------------|-------|--------|--------|
| 000000 – 00007F | 1 | 2 | 4 |
| 000080 – 00009F | 2 | | |
| 0000A0 – 0003FF | | | |
| 000400 – 0007FF | | | |
| 000800 – 003FFF | 3 | 4 | |
| 004000 – 00FFFF | 4 | | |
| 010000 – 03FFFF | | 4 | |
| 040000 – 10FFFF | | | |

5.5.e. Detección de la codificación de caracteres

Para leer adecuadamente un fichero de texto es imprescindible conocer la codificación de caracteres utilizada. Sucede a menudo que no conocemos exactamente en qué codificación de caracteres está el archivo que queremos leer y por este motivo existe una serie de algoritmos que intentan detectar la codificación automáticamente. Estos algoritmos funcionan de una manera heurística, a partir de análisis estadísticos de varios textos en diferentes lenguas y codificaciones de caracteres. Normalmente se trabaja con trigramas de caracteres de una manera similar a la que se hace para la detección automática de lengua¹.

Cuando el archivo está en Unicode UTF-8, la detección automática de la codificación de caracteres suele funcionar muy bien. Esto se explica por el gran porcentaje de secuencias de bytes inválidas en UTF-8, por lo que un archivo escrito en otra codificación muy difícilmente se detectará como UTF-8. La detección de la codificación UTF-16 también es bastante fiable dado el alto número de carácter de línea nueva (U + 000A) y espacios (U + 0020) que se encuentran cuando se dividen los datos en palabras de 16 bits (2 bytes). Cuando trabajamos con codificaciones ISO-8859 la detección de la codificación exacta resulta más compleja, ya que todas las codificaciones de esta familia comparten la parte baja de la tabla (que coincide totalmente con el ASCII).

Para evitar estos problemas, muchos formatos, como por ejemplo el HTML permiten especificar en qué codificación de caracteres está escrito el archivo.

¹ Mirar la sección *Para ampliar conocimientos*

```
<html>
<meta charset="UTF-8">
<body>
...
</body>
</html>
```

A continuación veremos un par de herramientas que nos permitirán detectar la codificación de caracteres de nuestros archivos.

Chardet

Chardet (<https://pypi.python.org/pypi/chardet>) es un paquete para el lenguaje de programación Python que permite detectar las siguientes codificaciones de caracteres:

- ASCII, UTF-8, UTF-16 (2 variants), UTF-32 (4 variants)
- Big5, GB2312, EUC-TW, HZ-GB-2 312, ISO-2022-CN (Traditional and Simplified Chinese)
- EUC-JP, SHIFT_JIS, ISO-2022-JP (Japanese)
- EUC-KR, ISO-2022-KR (Korean)
- KOI8-R, MacCyrillic, IBM855, IBM866, ISO-8859-5, windows-1251 (Cyrillic)
- ISO-8859-2, windows-1250 (Hungarian)
- ISO-8859-5, windows-1251 (Bulgarian)
- windows-1252 (English)
- ISO-8859-7, windows-1253 (Greek)
- ISO-8859-8, windows-1255 (Visual and Logical Hebrew)
- TIS-620 (tailandés)

Funciona bajo línea de comandos:

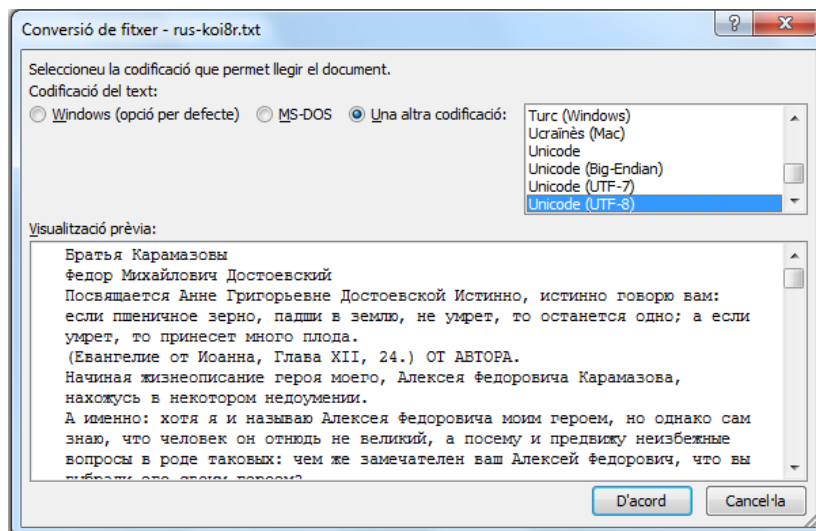
```
chardet doc4.txt
doc4.txt: KOI8-R (confidence: 1.00)
```

También puede detectar la codificación de un conjunto de ficheros:

```
chardet *
doc10.txt: UTF-16BE (confidence: 1.00)
doc1.txt: ISO-2022-JP (confidence: 0.99)
doc2.txt: UTF-16BE (confidence: 1.00)
doc3.txt: ISO-8859-2 (confidence: 0.87)
doc4.txt: KOI8-R (confidence: 1.00)
doc5.txt: ISO-2022-KR (confidence: 0.99)
doc6.txt: GB2312 (confidence: 0.99)
doc7.txt: UTF-8 (confidence: 1.00)
doc8.txt: UTF-8 (confidence: 1.00)
doc9.txt: ascii (confidence: 1.00)
```

Microsoft Word

Microsoft Word dispone de un buen algoritmo de detección de codificaciones de caracteres. Lo único que tenemos que hacer es ir a *File > Open* y seleccionar como formato *Encoded Text*. Entonces abrir el archivo de lo que queremos detectar la codificación y aparecerá una pantalla como la siguiente:



En esta pantalla nos indica la codificación de caracteres más probable y nos muestra un fragmento de texto para que podamos ver si la detección es correcta. En la mayoría de los casos lo es, pero si no es así, podemos seleccionar otra y veremos cómo se visualizaría el texto. Una vez seleccionamos la codificación correcta podremos abrir el archivo.

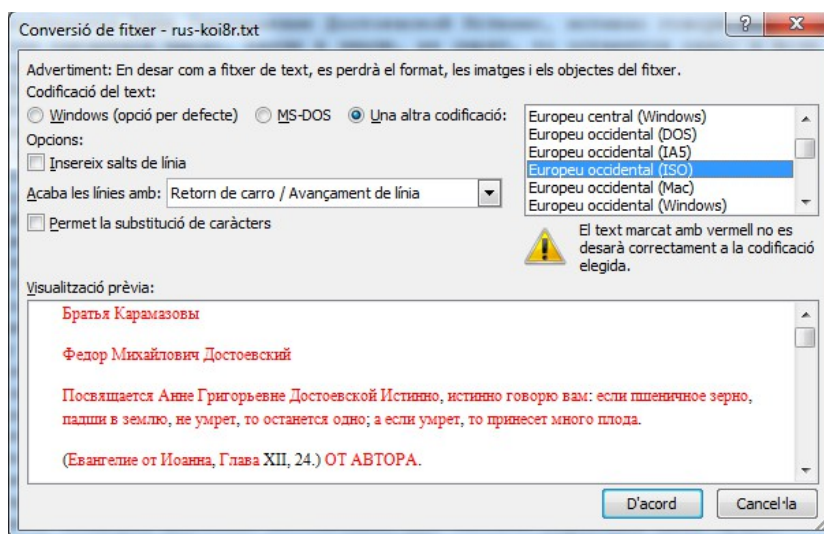
5.5.f. Cambio de la codificación de caracteres

En algunas situaciones es posible que necesitamos cambiar la codificación de caracteres de un determinado archivo. Si hemos sido capaces de detectarla (ver apartado anterior) la manera más sencilla de convertirla es usar un editor de textos con un buen apoyo en cuanto a la codificación de caracteres. Entonces, lo único que tenemos que hacer es abrir el archivo seleccionando la codificación adecuada y una vez abierto guardarlo haciendo *Save as* e indicando la codificación deseada.

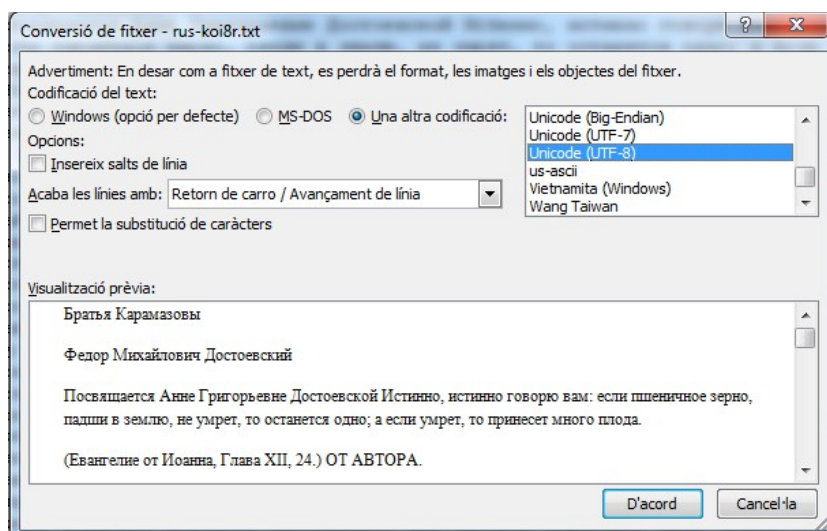
Veremos ahora cómo hacer esta operación en Microsoft Word y luego veremos un programa específico para hacer estas conversiones: el *iconv*.

Microsoft Word

Una vez abierto el archivo en la codificación adecuada (que muy probablemente habrá detectado automáticamente), lo único que tenemos que hacer es *Save as* y seleccionar otra codificación. La ventaja de utilizar esta aplicación es que es capaz de avisar si estamos seleccionando una codificación que no es capaz de guardar correctamente el archivo. Siguiendo el ejemplo anterior, una vez hemos abierto el archivo que está en codificación KOI8-R, si ahora lo intentamos guardar en codificación ISO-8859-1, en la pantalla nos mostrará los caracteres que no es capaz de guardar.



Por lo tanto, será necesario seleccionar una codifiación que sea capaz de guardar el archivo, por ejemplo Unicode UTF-8. En este caso podemos observar que no aparece ningún carácter en rojo.



Una vez guardado, el archiv tendrá la nueva codificación.

iconv

Es un programa y un conjunto de librerías para varios lenguajes de programación que permiten cambiar un archivo de una codificación a otra. Está disponible para Linux y Mac y también para Windows bajo Cygwin (<http://cygwin.com/>) y GNUWin32 (<http://gnuwin32.sourceforge.net/>).

Su funcionamiento es muy sencillo:

```
iconv -f koi8-r -t utf-8 < doc4.txt > doc4b.txt
```

Ahora el documento doc4b.txt estará en codificación UTF-8, y lo podemos verificar haciendo:

```
chardet doc4b.txt
doc4b.txt: utf-8 (confidence: 0.99)
```

5.6. La representación de la información no textual

En la unidad anterior hemos aprendido cómo se codifica la información básicamente textual (incluyendo saltos de línea y algunos otros caracteres de control) en un documento de texto. Los documentos con los que tendremos que trabajar contienen mucha más información, de carácter no textual, como pueden ser cuestiones de formato (negritas, tipo de letra, colores, etc.) o bien referencias a otros objetos (como pueden ser imágenes o gráficos). Toda esta información se codifica también de diferentes maneras. En esta unidad estudiaremos las maneras más habituales de codificar este tipo de información en los documentos.

El traductor a menudo han de traducir, además, ficheros que no son documentos en sentido estricto: páginas web, bases de datos, código de programas o archivos de imágenes.

Este tema no puede organizarse como una enumeración de los formatos de archivo más habituales, ya que la lista sería interminable. Lo que procuraremos será exponer algunas ideas básicas y estrategias para enfrentarnos a formatos de archivo desconocidos.

5.6.a. Nombres de archivo y extensiones. Relación con el formato y la aplicación

Suele haber una relación entre el formato del archivo y la extensión pero esta relación no es inequívoca, es decir, una misma extensión puede emplearse para diferentes tipos de archivos.

Recuerda que la *extensión* de un archivo viene dada por los caracteres que vienen detrás del punto (.) de su nombre. La extensión suele tener dos o tres letras. Por ejemplo, en el archivo:

```
resumen.txt
```

el nombre del archivo es "resumen" y la extensión "txt". En el siguiente archivo:

```
resumen de resultados de ventas.doc
```

el nombre es "resumen de resultados de ventas" y la extensión es "doc".

En algunos sistemas operativos por defecto se oculta la extensión de los tipos de archivos conocidos. Esto quiere decir, por ejemplo, que en los ejemplos anteriores, en nuestro explorador de archivos veríamos sólo "resumen" y "resumen de resultados de ventas", ya que las extensiones txt y doc muy probablemente serían conocidas para nuestro sistema.

El hecho de que nuestro sistema conozca una extensión o no viene dada por los programas que tenemos instalados. Por ejemplo, si no tenemos instalado el Open Office o el Libre Office en nuestro sistema, probablemente el sistema no conozca que la extensión "odt" corresponde a los documentos de este procesador de textos. Así, si tenemos activada la opción de ocultar las extensiones conocidas en nuestro explorador de archivos, muy probablemente este mostrará las extensiones de los archivos "odt". Si instalamos el Libre Office u Open Office, en el momento de la instalación se indicará al sistema operativo que esta extensión está asociada a esta aplicación y por tanto la extensión pasará a ser conocida para el sistema, y se ocultará si tenemos activada la opción correspondiente.

En muchas situaciones puede ser recomendable desactivar la opción de ocultar la extensión de tipos de archivos conocidos de nuestro sistema operativo.

El sistema operativo, pues, tiene asociadas algunas aplicaciones a algunas extensiones. Para estas extensiones, si hacemos doble clic en un archivo desde el explorador de archivos, el sistema operativo pondrá en marcha la aplicación asociada y abrirá el archivo con esta aplicación. Si el sistema operativo no tiene asociada ninguna aplicación a una extensión determinada, si hacemos doble clic en un archivo de este tipo el sistema no sabrá qué hacer y no podrá abrirlo.

También es posible indicar manualmente qué programa está asociado a cada extensión.

El hecho de que un archivo no se abra haciendo doble clic en el explorador de archivos no quiere decir que no podamos abrirlo. Una situación habitual es recibir un archivo de texto con una extensión no estándar (por ejemplo un archivo `resumen.rsq`). Si quien nos envía el archivo nos dice que es de texto, podremos ir a un editor de textos y abrirlo desde el menú *File* o similar. Si no nos dicen nada de qué es el archivo, podemos llevar a cabo las siguientes acciones:

- Buscar en alguna base de datos de extensiones si aparece. Algunas webs que ofrecen estas búsquedas son <http://filext.com/> o <http://www.fileinfo.com/>
- Suponer que es un archivo de texto e intentar abrirlo directamente con un editor de textos
- Si trabajamos con Linux o Mac podemos utilizar la instrucción *file* que nos intentará determinar el tipo de archivo. Si hacemos:

```
file resum.rsq
```

Obtendremos la información sobre el tipo de archivo:

```
resum.rsq: UTF-8 Unicode text
```

En este caso tendremos que abrir este fichero con algún editor de textos. Si recibiéramos menudo archivos de este tipo podríamos asociar esta extensión en el editor de textos de nuestro sistema operativo, por lo que a partir de este momento al hacer doble clic sobre archivos "rsq" se abrirían con el editor seleccionado.

5.6.b. El formato HTML

El *Hyper Text Markup Language* (HTML) es un lenguaje de marcado derivado del SGML (*Standard Generalized Markup Language*), diseñado para visualizar textos y relacionarlos en forma de hipertexto.

En HTML se utilizan una serie de etiquetas que definen cómo se visualizará el texto en un navegador. Las etiquetas más habituales son (fuente Wikipedia

http://ca.wikipedia.org/wiki/Hyper_Text_Markup_Language):

- **<html>** Es la etiqueta raíz de cualquier documento HTML.
- **<head>**: Define la cabecera del documento HTML.
- **<body>**: Define el cuerpo del documento. Esta es la parte del documento HTML que se muestra en el navegador.

Dentro de la cabecera **<HEAD>** podemos encontrar:

- **<title>**: Permite definir el título de la página. En navegadores gráficos el contenido del *title* aparece en la barra del título encima de la ventana.
- **<meta>**: Permite definir metainformación del documento tales como el autor, la fecha de realización, la codificación del documento (UTF, ISO, etc), las palabras clave y la descripción del mismo
- **<LINK>**: Permite definir metadatos complementarios a las del meta tales como el documento anterior, el siguiente, el capítulo al que pertenece el documento, la página, glosario, etc.

Dentro del cuerpo **<BODY>** podemos encontrar:

- **<a>**: Etiqueta ancla. Crea un enlace a otro documento o a otra zona del mismo, según los atributos.
- **<h1>**, **<h2>**, ... **<h6>**: cabeceras o títulos del documento, suelen distinguirse por tamaño.
- **<div>**: División estructural de la página.
- **<p>**: Párrafo.
- **
**: Salto de línea.
- **<table>**: Indica el comienzo de una tabla, después se tendrá que definir las filas con **<tr>** y las celdas dentro de las filas con **<td>**.
- ****: Lista desordenada (sin numerar). Los ítems se definen con ****.
- ****: Lista ordenada (numerada). Los ítems se definen con ****.
- **<dl>**: Lista de definición. Hay dos tipos de ítem; el **dt** y el **dd**.
 - **<dt>**: Término a definir.
 - **<dd>**: Definición del término.

Por regla general las etiquetas que se abren deben cerrarse (aunque la mayoría de navegadores permiten omitir el cierre de algunas etiquetas). Vemos ahora como ejemplo un documento HTML muy sencillo:

```
<html>
<body>
<p>Así se muestran <b>también</b> los caracteres <i>acentuados</i>.</p>
</body>
</html>
```

En un navegador de Internet este html se visualizaría de la siguiente manera:

Así se muestran **también** los caracteres *acentuados*.

Nótese que el "í" se ha expresado como í y le "é" como é. Estas combinaciones de caracteres para expresar caracteres especiales no incluidos en el alfabeto latino básico se denominan *entidades de html*.

Estas entidades también se pueden expresar de forma numérica:

```
<html>
<body>
<p>Así se muestran <b>también</b> los caracteres <i>acentuados</i>.</p>
</body>
</html>
```

En el Anexo I de este capítulo presentamos una lista muy completa de las entidades de html.

Sólo es imprescindible utilizar entidades de html para representar los caracteres reservados, que son: <, > y &. Si queremos escribir en html algo como:

Si a > b & b < c y lo hacemos:

```
<html>
<body>
<p>Si a > b & b < c</p>
</body>
</html>
```

El navegador podría confundirse (aunque la mayoría de navegadores modernos son capaces de representar este documento sin problemas). Estrictamente, deberíamos representar los documento de la siguiente manera:

```
<html>
<body>
<p>Si a &gt; b &amp; b &lt; c</p>
</body>
</html>
```

Nuestro ejemplo anterior lo podemos escribir sin problemas utilizando los caracteres acentuados normales:

```
<html>
<body>
<p>Así se muestran <b>también</b> los caracteres <i>acentuados</i>.</p>
</body>
</html>
```

En este caso, si el html no especifica la codificación de caracteres, dependiendo de la configuración de nuestro navegador podemos ver el documento de manera incorrecta:

AsÃ se muestran **tambiÃ** los caracteres *acentuados*.

En la configuración de nuestro navegador podemos indicar otra codificación de caracteres para hacer que el documento se visualice correctamente. Ahora bien, para evitar estos problemas, es útil indicar la codificación de caracteres utilizada dentro del propio html, por lo que el navegador utilizará esta información para visualizar el documento de manera correcta.

```
<html>
<meta charset="UTF-8">
<body>
<p>Así se muestran <b>también</b> los caracteres <i>acentuados</i>.</p>
</body>
</html>
```

Y ahora el navegador utilizará esta información y visualizará correctamente los caracteres acentuados:

Así se muestran **también** los caracteres *acentuados*.

5.6.c. *El XHTML*

El XHTML es como el HTML pero escrito como un XML (veremos con detalle el XML un poco más adelante en este mismo capítulo). El XHTML se puede definir como una versión de HTML más estricta y limpia. En HTML se ha permitido escribir documentos con etiquetas que se abren y no se cierran, problemas de anidamiento de etiquetas, etc. y los navegadores han sido capaces de visualizar las páginas igualmente. Es decir, se ha primado la robustez a la corrección en la sintaxis. En XHTML se verifica que el documento sea correcto de manera estricta y si no lo es el navegador impide su visualización.

Volviendo a nuestro ejemplo anterior, pero introducimos un error:

```
<html>
<meta charset="UTF-8">
<body>
<p>Así se muestran <b>también</b> los caracteres <i>acentuados</i>.</p>
</body>
</html>
```

(Fíjese que la etiqueta `<p>` no se cierra). Si guardamos este archivo como ejemplo.htm (es decir, un html normal) y lo intentamos visualizar en un navegador, lo visualizaremos sin problemas. En cambio, si lo guardamos como ejemplo.xhtml (es decir, como XHTML), al intentar abrirlo en un navegador, nos mostrará un mensaje como el siguiente:

This page contains the following errors:

error on line 5 at column 8: Opening and ending tag mismatch: p line 0 and body

Below is a rendering of the page up to the first error.

Así se muestran también los caracteres acentuados

5.6.d. Open Document

El ODF (*Open Document Format for Office Applications - Documento Abierto para Aplicaciones Informáticas*) es un formato basado en XML para la representación de documentos, hojas de cálculo, gráficos y presentaciones. El estándar fue desarrollado por un comité técnico del consorcio OASIS (*Organisation for the Advancement of Structured Information Standards*). *Se basa en la especificación XML de OpenOffice.org de Sun Microsystems.*

Además de ser un estándar OASIS, la versión 1.1 es también un estándar internacional ISO/IEC 26300:2006/Amd 1:2012 - Open Document Format for Office Applications (OpenDocument) v1.1.).

Las extensiones asociadas a Open Document son:

- Texto: .odt
- Hoja de cálculo: .ods
- Presentación: .odp
- Dibujo: .odg
- Gráfico: .odc
- Fórmula matemática: .odf
- Base de datos: .ODB
- Imagen: .odi
- Documento maestro: .odm

Y las asociadas a las plantillas son:

- Texto: .ott
- Hoja de cálculo: .ots
- Presentación: .otp
- Dibujo: .otg

Los archivos Open Document son archivos comprimidos ZIP que contienen varios archivos y directorios:

- Directorios
 - META-INF
 - Thumbnails
 - Pictures
 - Configurations2
- Archivos XML
 - content.xml
 - meta.xml
 - settings.xml
 - styles.xml
- Otros archivos
 - mimetype
 - layout-cache

Para ver por dentro el contenido de estos archivos hemos creado un documento que contiene:

Esto es un **documento** de *ejemplo*?

Y lo hemos guardado como documento.odt. Cambiamos la extensión a documento.zip y lo descomprimimos.

El archivo **content.xml** almacena el contenido real del documento, exceptuando los datos binarios como imágenes. Si abrimos este archivo con un editor de textos obtendremos algo similar a lo siguiente (aquí hemos simplificado el contenido):

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content
xmlns:office="urn:oasis:names:tc:opendocument:xmlns:office:1.0" [...]
<style:style style:name="T1" style:family="text"><style:text-properties
fo:font-weight="bold" style:font-weight-asian="bold" style:font-weight-
complex="bold"/></style:style><style:style style:name="T2"
style:family="text"><style:text-properties fo:font-style="italic" style:font-
style-asian="italic" style:font-style-complex="italic"/> [...]
</text:sequence-decls><text:h text:style-name="Standard" text:outline-
level="10">Esto es un <text:span text:style-name="T1">documento</text:span>
de <text:span text:style-name="T2">ejemplo</text:span>?
</text:h></office:text></office:body></office:document-content>
```

Buena parte de la información sobre los estilos de formato y disposición del documento se almacena en el archivo **styles.xml**. No toda la información sobre los estilos almacena en este archivo, también hay información sobre estilos a **content.xml**. Fíjate cómo se definen los estilos T1 y T2 en el archivo **content.xml** anterior. A continuación mostramos algunos fragmentos del archivo **styles.xml**:

```
[...]
<style:font-face style:name="Courier 10 Pitch" svg:font-family="&apos;Courier
10 Pitch&apos;" style:font-pitch="fixed"/>
[...]
<style:style style:name="Text_20_body" style:display-name="Text body"
style:family="paragraph" style:parent-style-name="Standard"
style:class="text"><style:paragraph-properties fo:margin-top="0in" fo:margin-
bottom="0.0835in"/></style:style>
[...]
<style:list-level-label-alignment text:label-followed-by="listtab" text:list-
tab-stop-position="0.7in" fo:text-indent="-0.7in" fo:margin-left="0.7in"/>
```

El archivo **meta.xml** contiene los metadatos del documento, como el autor, el usuario que hizo la última modificación, la fecha de creación y de la última modificación, así como algunas estadísticas del documento como el número de tablas, imágenes, páginas, párrafos, palabras, etc. A continuación podemos observar el contenido de este archivo:

```
<?xml version="1.0" encoding="UTF-8"?>
[...]
<office:meta>
<meta:initial-creator>Antoni Oliver</meta:initial-creator>
<meta:creation-date>2014-08-14T18:20:54</meta:creation-date>
<dc:date>2014-08-14T18:22:52</dc:date>
<dc:creator>Antoni Oliver</dc:creator>
<meta:editing-duration>POD</meta:editing-duration>
<meta:editing-cycles>1</meta:editing-cycles>
<meta:document-statistic meta:table-count="0" meta:image-count="0"
meta:object-count="0" meta:page-count="1" meta:paragraph-count="1" meta:word-
count="5" meta:character-count="30" meta:non-whitespace-character-
count="26"/>
<meta:generator>LibreOffice/3.5$Linux_x86 LibreOffice_project/350ml$Build-
2</meta:generator></office:meta></office:document-meta>
```

El archivo **settings.xml** contiene información que no hace referencia ni a contenido ni a disposición, sino información que hace referencia a aspectos de la visualización del documento dentro de la aplicación en

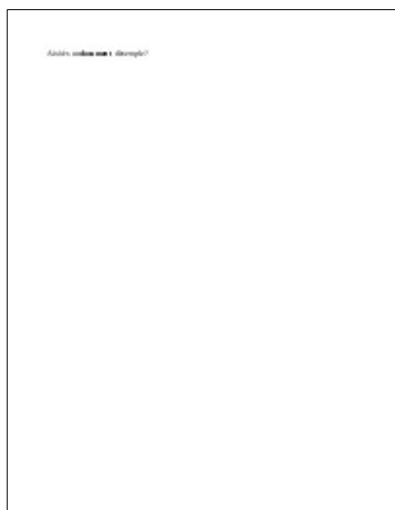
el momento en que se abre el documento. Esta información puede hacer referencia a las áreas de visualización, la posición del cursor o el factor de zoom. A continuación podemos observar un fragmento de este archivo:

```
[...]
<config:config-item config:name="ViewAreaTop"
config:type="int">0</config:config-item>
<config:config-item config:name="ViewAreaLeft"
config:type="int">0</config:config-item>
<config:config-item config:name="ViewAreaWidth"
config:type="int">21527</config:config-item>
<config:config-item config:name="ViewAreaHeight"
config:type="int">9289</config:config-item>
<config:config-item config:name="ShowRedlineChanges"
config:type="boolean">true</config:config-item>
<config:config-item config:name="InBrowseMode"
config:type="boolean">false</config:config-item>
[...]
<config:config-item config:name="ZoomFactor"
config:type="short">150</config:config-item>
<config:config-item config:name="IsSelectedFrame" config:type="boolean">false
[...]
```

El archivo **mimetype** tiene una sola línea y contiene información sobre el tipo de archivo. Esto hace que de hecho la extensión del archivo sea irrelevante y sólo sirva para que el usuario pueda identificar más fácilmente el tipo de documento. En nuestro ejemplo el archivo mimetype contiene la siguiente línea:

```
application/vnd.oasis.opendocument.text
```

La carpeta **Thumbnails** contiene una imagen en miniatura de la primera página del documento y se genera por defecto cuando se guarda el archivo. La imagen está en formato png y tiene un tamaño de 128x128 píxeles. En nuestro ejemplo tendría el siguiente aspecto:



La carpeta **META-INF** contiene un archivo **manifest.xml** que contiene información sobre los archivos contenidos en el fichero comprimido OpenDocument. A continuación vemos el contenido de este archivo correspondiente a nuestro ejemplo:


```
<?xml version="1.0" encoding="UTF-8"?>
<manifest:manifest
xmlns:manifest="urn:oasis:names:tc:opendocument:xmlns:manifest:1.0"
manifest:version="1.2">
<manifest:file-entry manifest:full-path="/" manifest:version="1.2"
manifest:media-type="application/vnd.oasis.opendocument.text"/>
<manifest:file-entry manifest:full-path="meta.xml" manifest:media-
type="text/xml"/>
<manifest:file-entry manifest:full-path="settings.xml" manifest:media-
type="text/xml"/>
<manifest:file-entry manifest:full-path="content.xml" manifest:media-
type="text/xml"/>
<manifest:file-entry manifest:full-path="Thumbnails/thumbnail.png"
manifest:media-type="image/png"/>
<manifest:file-entry manifest:full-path="manifest.rdf" manifest:media-
type="application/rdf+xml"/>
<manifest:file-entry manifest:full-
path="Configurations2/accelerator/current.xml" manifest:media-type=""/>
<manifest:file-entry manifest:full-path="Configurations2/" manifest:media-
type="application/vnd.sun.xml.ui.configuration"/>
<manifest:file-entry manifest:full-path="styles.xml" manifest:media-
type="text/xml"/>
</manifest:manifest>
```

La carpeta **Pictures** contiene todas las imágenes del documento. Estas imágenes se almacenan en su formato original, a excepción de las imágenes en mapa de bits que se transforman en PNG por cuestión de espacio. Esta carpeta sólo aparece si el documento contiene imágenes. La información sobre la imagen y como aparece en el documento aparecerá en el archivo content.xml. Vemos aquí un ejemplo:

```
<draw:frame draw
:style-name="fr1" draw:name="graphics1" text:anchor-type="paragraph"
svg:width=
"6.2575in" svg:height="1.8043in" draw:z-index="0"><draw:image
xlink:href="Pictu
res/100000000000025F000000AFF9CA7EDD.png" xlink:type="simple"
xlink:show="embed
" xlink:actuate="onLoad"/></draw:frame>
```

5.6.e. Los formatos de documentos de Microsoft Word

Microsoft Word DOC (.doc)

El formato **doc** correspondiente al Microsoft Word (versiones del 97 al 2003) es un formato binario. Aunque Microsoft publicó las especificaciones de este formato, siempre ha habido la queja de que estas especificaciones no son completas. Por este motivo es difícil abrir documentos en este formato en procesadores de textos diferentes de Microsoft Word. Aunque procesadores de textos libres como el LibreOffice y el Open Office pueden abrir este formato, no se puede garantizar una compatibilidad al 100% y es posible que algunas características de los documentos se pierdan.

Las herramientas de Traducción Asistida que incluyen la posibilidad de importar documentos Word tipo doc normalmente requieren tener Microsoft Word instalado en el ordenador. En realidad la herramienta no abre el documento word, sino que hace abrir el documento en Microsoft Word y se comunica con esta aplicación para obtener los segmentos traducibles. Esta comunicación se realiza mediante una API (*Application Programming Interface*). Una vez hecha la traducción con la herramienta de traducción asistida, en el momento de crear el documento traducido se vuelve a establecer una comunicación con el Microsoft Word que reemplaza los segmentos originales por traducidos y crea de esta manera un documento traducido con el mismo formato que el original.

Microsoft Word 2003 XML (.xml)

Microsoft introdujo un formato XML en el Office. A diferencia del Open Document estos documentos están formado por un único archivo XML. Si continuamos con el mismo documento de ejemplo, en este formato tendría el siguiente aspecto (mostramos únicamente un fragmento del documento):

```
<w:pStyle w:val="Standard"/></w:pPr><w:r><w:t>Esto es un
</w:t></w:r><w:r><w:rPr><w:rStyle
w:val="T1"/></w:rPr><w:t>documento</w:t></w:r><w:r><w:t>
de</w:t></w:r><w:r><w:rPr><w:rStyle
w:val="T2"/></w:rPr><w:t>ejemplo</w:t></w:r><w:r><w:t>?</w:t></w:r></w:p>....
```

Si el documento contiene imágenes u otros objetos, estos quedan también representados en el documento XML, por ejemplo:

```
<w:pict><w:binData
w:name="wordml://graphics1">iVBORw0KGgoAAAANSUUhEUgAAAEAAAC5CAIAAACH1v1PAAAAA
3NCSVQICAjb4U/gAAAgAE1EQVR4nO3dd1gU1/o48Bd26V2aCAKKorQA0hQRUKQGQUTEgkZNRjUaNd
EYrz0mQSVCNBbUWLAXIBIQGxE1iiJNaTYEBekgvpP7[... ]u+ADxAAAAAE1FTkSuQmCC</w:binDa
ta>
```

Microsoft Word 2007/2010 XML (.docx)

Posteriormente, a partir de la versión 2007, Microsoft introdujo un formato basado en XML consistente en varios archivos en un archivo zip, llamado Office Open XML (y también conocido de manera informal como OOXML o OpenXML). La idea básica es similar a la que hemos explicado un poco más arriba sobre el formato ODF. El formato alcanzó la calificación de estándar primero por la Ecma (as ECMA-376) y en versiones posteriores por ISO e IEC (como ISO/IEC 29500).

A partir de la versión de Office 2007 este formato es el formato por defecto en esta aplicación ofimática.

La estructura de los archivos docx es la siguiente:

- Archivo [Content_Types] .xml
- Carpeta docProps
- Carpeta _rels
- Carpeta word

No entraremos en detalles, sólo diremos que el contenido del documento está en un archivo document.xml que está dentro de la carpeta word. Mostramos el correspondiente al archivo de ejemplo:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:document xmlns:o="urn:schemas-microsoft-com:office:office"
xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
xmlns:v="urn:schemas-microsoft-com:vml"
xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
xmlns:w10="urn:schemas-microsoft-com:office:word"
xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDraw
ing">
<w:body>
<w:p>
<w:pPr>
<w:pStyle w:val="style0"/>
<w:numPr>
<w:ilvl
w:val="8"/>
<w:numId w:val="1"/>
</w:numPr>
<w:spacing w:line="100"
w:lineRule="atLeast"/>
</w:pPr>
<w:r>
<w:rPr>
<w:lang w:val="es-ES"/>
</w:rPr>
<w:t
xml:space="preserve">
Esto es un </w:t>
</w:r>
<w:r>
<w:rPr>
<w:b/>
<w:bCs/>
<w:lang
w:val="es-ES"/>
</w:rPr>
<w:t>
documento</w:t>
</w:r>
<w:r>
<w:rPr>
<w:lang
w:val="es-ES"/>
</w:rPr>
<w:t
xml:space="preserve">
de</w:t>
</w:r>
<w:r>
<w:rPr>
<w:i/>
<w:iCs/>
<w:lang w:val="es-
ES"/>
</w:rPr>
<w:t>
ejemplo</w:t>
</w:r>
<w:r>
<w:rPr>
<w:lang w:val="es-
ES"/>
</w:rPr>
<w:t>
?</w:t>
</w:r>
</w:p>
<w:sectPr>
<w:type
w:val="nextPage"/>
<w:pgSz w:h="15840" w:w="12240"/>
<w:pgMar w:bottom="1134"
w:footer="0" w:gutter="0" w:header="0" w:left="1134" w:right="1134"
w:top="1134"/>
<w:pgNumType w:fmt="decimal"/>
<w:formProt
w:val="false"/>
<w:textDirection
w:val="lrTb"/>
</w:sectPr>
</w:body>
</w:document>
```

Si el documento contiene imágenes, éstas se almacenan en una subcarpeta de la carpeta word, llamada media.

5.6.f. El formato LaTeX

LaTeX es un lenguaje de marcado de documentos y un sistema de edición de documentos que se utiliza mucho para la publicación de documentos y libros científicos. Los documentos LaTeX se pueden escribir en cualquier editor de textos, aunque hay algunos específicos que contienen algunas funciones de ayuda (como por ejemplo Kile (<http://kile.sourceforge.net/>) o TexMaker (<http://www.xmlmath.net/texmaker/>)). Los ficheros LaTeX suelen tener la extensión tex. Además del editor se necesita tener instalados los programas y macros que permiten la transformación de los archivos tex en otros formatos como el dvi, ps o pdf. Estas macros se pueden instalar fácilmente en sistemas operativos como Linux y Mac. Para Windows se puede usar el entorno MikTeX (<http://www.miktex.org/>) y que se puede configurar para funcionar de manera integrada en el editor TeXnicCenter (<http://www.texniccenter.org/>).

A continuación vemos un ejemplo de documento LaTeX mínimo:

```
\documentclass[12pt]{article}
\usepackage[utf8]{inputenc}
\begin{document}
Esto es un \bf{documento} de \emph{ejemplo}.
\end{document}
```

Para procesar este documento usaremos las siguientes instrucciones (si el documento se llama documento.tex):

```
latex documento.tex
```

Y se mostrará por pantalla:

```
This is pdfTeX, Version 3.1415926-1.40.10 (TeX Live 2009/Debian)
entering extended mode
(./latex.tex
LaTeX2e <2009/09/24>
Babel <v3.81> and hyphenation patterns for english, usenglishmax, dumylang,
nohyphenation, loaded.
(/usr/share/texmf-texlive/tex/latex/base/article.cls
Document Class: article 2007/10/19 v1.4h Standard LaTeX document class
(/usr/share/texmf-texlive/tex/latex/base/size12.clo)
(/usr/share/texmf-texlive/tex/latex/base/inputenc.sty
(/usr/share/texmf-texlive/tex/latex/base/utf8.def
(/usr/share/texmf-texlive/tex/latex/base/tlenc.dfu)
(/usr/share/texmf-texlive/tex/latex/base/otlenc.dfu)
(/usr/share/texmf-texlive/tex/latex/base/omsenc.dfu))) (./latex.aux) [1]
(./latex.aux) )
Output written on documento.dvi (1 page, 364 bytes).
Transcript written on document.log.
```

Aquí se habrá creado un archivo xdvi que tendrá el siguiente aspecto:

Esto es un **documento** de *ejemplo*.

Si ahora queremos convertir este archivo xdvi en pdf utilizaremos las instrucciones:

```
dvips documento.dvi
ps2pdf documento.ps
```

Alternativamente, se puede convertir un documento LaTeX en pdf con una única instrucción:

```
pdflatex documento.tex
```

Muchos entornos gráficos permiten hacer estas operaciones con un solo clic en un botón.

El LaTeX se utiliza mucho en el mundo académico porque permite escribir fácilmente fórmulas matemáticas y porque se puede gestionar la bibliografía de manera muy fácil y eficiente con el entorno BibTeX

5.6.g. El formato DocBook

DocBook es un lenguaje de marcado semántico basado en XML que sirve para representar cualquier tipo de documento, tales como libros, manuales y artículos académicos. Del mismo modo que ocurría con LaTeX, los usuarios pueden crear el contenido de la publicación sin preocuparse de su presentación ni del formato final. Una vez terminado el documento en formato DocBook se pueden crear fácilmente los documentos finales en diversos formatos (entre ellos PDF, EPUB, HTML, XHTML, etc.) sin tener que hacer ningún cambio en el documento.

Las etiquetas de los documentos DocBook se pueden dividir en tres grandes categorías: estructurales, de bloque y de línea.

Algunas de las etiquetas estructurales son:

set: Es un conjunto de uno o más **book**. La ventaja de usar set es que se pueden utilizar los enlaces entre todos los libros

book: se estructuran de la siguiente manera:

```
book
  meta information
  chapter
    sect1
    sect1
  chapter
    sect1
  appendix
    sect1
  appendix
    sect1
  ...
  glossary
```

article: Se estructura de la siguiente manera:

```
article
  meta information
  sect1
  sect1
    sect2
  sect1
  ...
```

Las etiquetas de bloque son elementos tales como párrafos, listas, barras laterales, tablas y citas.

Las etiquetas de línea representan elementos como énfasis, hyper-enlaces, etc. que hacen que se aplique algún tipo de distinción tipográfica al texto, como cambios del tamaño de la fuente, fuente en cursiva o negrita, etc

A continuación podemos observar un fragmento de un libro en formato DocBook:

```
<!DOCTYPE book PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN"
    "http://www.oasis-open.org/docbook/xml/4.1.2/docbookx.dtd">
<book>
<title>THE ADVENTURES OF SHERLOCK HOLMES</title>
<bookinfo>
<author>
<firstname>Arthur Conan</firstname>
<surname>Doyle</surname>
</author>
</bookinfo>
```

```
<chapter>
<title>ADVENTURE I. A SCANDAL IN BOHEMIA</title>
<section>
<title>I.</title>
<para>To Sherlock Holmes she is always <emphasis>the</emphasis> woman. I have seldom heard him
mention her under any other name. In his eyes she eclipses and predominates the whole of her sex.
... And yet there was but one woman to him, and that woman was the late Irene Adler, of dubious
and questionable memory.</para>
...
</section>
...
</chapter>
...
</book>
```

Para editar documentos DocBook, como son documentos XML, sólo necesitamos un editor de textos. Podemos utilizar editores de textos de XML que nos facilitarán el trabajo de edición.

Dado que DocBook es un formato XML, podemos utilizar herramientas estándar para validar y procesar documentos DocBook y transformarlos en otros formatos. Se puede encontrar información detallada sobre cómo procesar documentos DocBook a Stayton (2007).

5.6.h. El formato PDF

PDF (acrónimo en inglés de *Portable Document Format*, Formato de Documento Portable) es un formato desarrollado por la empresa Adobe con la idea de que el documento se pueda visualizar exactamente igual con independencia del software, hardware o sistema operativo utilizado

Características

Las principales características del formato PDF son las siguientes:

- Es multiplataforma, se puede visualizar en los principales sistemas operativos como GNU / Linux, Windows o Mac, respetando el aspecto original.
- Puede guardar una combinación de texto, gráficos, imágenes e incluso audio.
- Es uno de los formatos más extendidos en Internet y es utilizado tanto por gobiernos como por empresas.
- Tiene la especificación abierta, permite incluso distribuir herramientas para crear, visualizar o modificar documentos en formato PDF como software libre.
- Puede cifrarse para proteger su contenido e incluso firmar electrónicamente.

Visualización

Para poder visualizar archivos PDF es necesario disponer de un software específico, pero que son gratuitos y algunos incluso de software libre. Podemos destacar los siguientes:

- Adobe Acrobat reader (<http://get.adobe.com/reader>): disponible para Windows, Linux y Mac, entre otros. Es gratuito, aunque no de software libre.

- Okular (<http://okular.kde.org/>): es de software libre y funciona bajo Linux, aunque también es posible instalarlo en Windows.
- XPDF (<http://www.foolabs.com/xpdf/>): es de software libre y funciona bajo Linux, Windows y Mac.

Creación

Hay muchas opciones para crear archivos PDF:

En Linux y Mac se incluyen utilidades para imprimir archivos en formato PDF. En Windows se pueden añadir impresoras virtuales, como por ejemplo PDF Creator (<http://sourceforge.net/projects/pdfcreator/>). LibreOffice y OpenOffice, así como las últimas versiones de Microsoft Office, permiten crear archivos PDF

Google Docs (<https://docs.google.com>) y Google Drive (<https://drive.google.com/>) permiten también crear y cargar archivos y luego guardarlos en PDF

Edición

Para poder editar y modificar archivos PDF es necesario disponer de un editor. Adobe ofrece soluciones propietarias y de pago (ya sea de compra o de suscripción mensual). Existen también algunos programas libres que permiten editar PDF, entre los que podemos destacar:

- PDFedit (<http://pdfedit.cz/en/index.html>): es una completa librería para la manipulación de documentos PDF. Dispone de una interfaz gráfica para facilitar su uso.
- pdftk (<https://www.pdfabs.com/tools/pdftk-the-pdf-toolkit/>): Dispone de versiones gratuita y de pago. La versión gratuita permite realizar diversas operaciones, tales como partir o juntar documentos pdf, hacer rotaciones, etc. Funciona bajo Windows y Linux.
- Nitro PDF (<http://www.nitropdf.com/>): es un programa propietario y que es de pago en su versión PRO. Ofrece un Reader gratuito pero que tiene más funcionalidades que un simple lector: creación de PDFs, conversión de PDF a texto, capturas de secciones del PDF, extraer las imágenes del PDF, etc.

Transformar PDFs a formatos editables

A menudo es necesario transformar un archivo PDF en un formato editable, tales como texto, ODF o word. Como traductor es posible que recibas archivos PDFs para traducir y que los quieras tratar con una herramienta de traducción asistida. Para ello será necesario convertir el PDF en un formato editable. Hay que tener en cuenta, sin embargo, que esta conversión no será siempre perfecta y que en muchos casos perderemos todo el formato del documento. Hay que recordar que los archivos PDF se generan siempre a partir de archivos editables. En caso de recibir para traducir un PDF, convendría preguntar al cliente si dispone del documento editable original, ya que seguramente será más fácil de tratar. En caso de que no sea posible disponer del documento en el formato original podremos transformarlo en texto de alguna de las siguientes maneras:

- Con algún programa de edición de PDFs, como los expuestos un poco más arriba.
- En algunos casos, será posible utilizar un programa de visualización y seleccionar el texto, copiarlo al portapapeles y pegarlo en un documento. Algunos documentos PDF, como explicaremos un poco más adelante, no permiten la selección de texto.
- En Linux disponemos de programas que funcionan bajo terminal que permiten hacer la conversión, como pdftotext o pdf2txt. Para Windows disponemos de algunas aplicaciones como pdf2textpilot (<http://sourceforge.net/projects/pdf2textpilot/>). Xpdf (<http://www.foolabs.com/xpdf/>), disponible para Linux, Mac y Windows, dispone, entre otras funcionalidades de una implementación de pdftotext.
- Hay que tener en cuenta que los archivos PDF creados a partir del escaneado de documentos en papel que contienen texto no tienen la misma estructura que el mismo PDF que se hubiera creado directamente desde la aplicación correspondiente. El documento PDF proveniente del escaneado

internamente contiene una imagen del documento, sin ninguna información respecto al texto. Para poder pasar a texto este documento será necesario utilizar técnicas de OCR (*Optical Character Recognition*). Algunos editores de PDF ya implementan estas técnicas para poder tratar este tipo de documento. Si no es así, podremos utilizar algún programa específico de OCR (como Tesseract OCR -

- <http://code.google.com/p/tesseract-ocr/>, que tiene licencia libre). Algunas versiones de Microsoft Windows disponen de la funcionalidad de OCR, que se encuentra en el menú de accesorios. También es posible utilizar algún servicio gratuito de OCR on-line, como por ejemplo <http://www.free-online-ocr.com/> o <http://www.onlineocr.net/>.
- Google Drive (<https://drive.google.com/>) permite subir archivos PDF. En el momento de subirlos permite seleccionar si deseamos convertir el documento PDF en un documento de Drive (que es editable). Si seleccionamos esta opción se llevará a cabo la conversión, y si el PDF proviene de un documento escaneado, se llevará a cabo un OCR.

Ten en cuenta que si utilizas un OCR, ya sea un programa, como un servicio on-line, como el propio Google Drive, conviene indicar la lengua del documento. Esta información es importante, ya que la precisión del OCR aumentará, ya que en caso de dudas, elegirá una palabra de la lengua.

5.7. XML

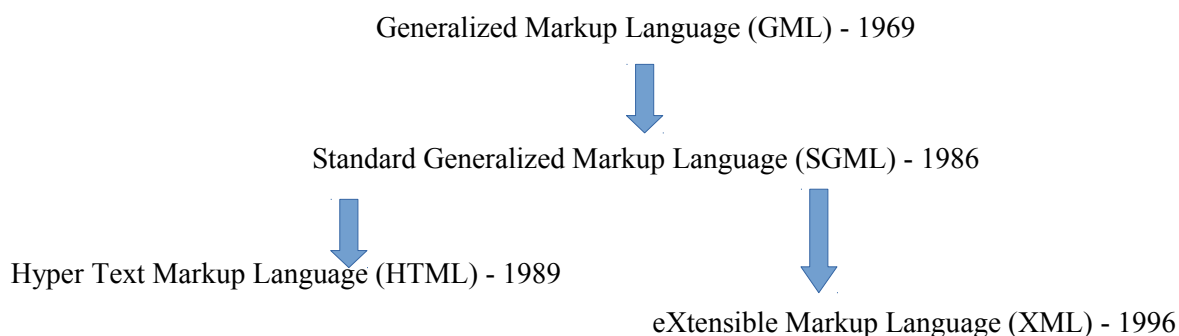
5.7.a. Introducción

El *eXtensible Markup Language* (XML) es un lenguaje de marcado que define una serie de reglas para representar información estructurada de una manera que es fácilmente legible tanto para los humanos como para las máquinas. El XML no es, por tanto, un lenguaje en particular, sino una manera de definir lenguajes para diferentes necesidades: representar documentos, bases de datos, en servicios web como contenedor de la información, etc. XML proporciona un modo sencillo para representar y transmitir información y existe toda una serie de tecnologías asociadas que permiten un tratamiento sencillo y eficiente de estos tipos de documentos.

En esta sección haremos una introducción general, centrándonos en aquellos aspectos que pueden ser más interesantes para un traductor. Quien quiera una introducción más completa y general puede consultar el tutorial de XML de W3Schools en <http://www.w3schools.com/xml/>

Lo primero que debemos tener en cuenta es que el XML es un formato de texto, y que por tanto, los archivos XML se pueden abrir para visualizar y modificar en cualquier editor de textos. Si visualizamos un archivo XML la primera impresión que seguramente tendremos es que se parece mucho al HTML. Esto tiene una explicación, el HTML es un lenguaje que se derivó del SGML (*Standard Generalized Markup Language*) y de hecho el XML es una simplificación de SGML, y por tanto, se puede considerar también un lenguaje derivado. Sin embargo hay que recordar dos cosas: que el XML no es un sustituto del HTML y que ambos lenguajes se diseñaron con dos objetivos bien diferentes: el XML está diseñado para describir datos y HTML está diseñado para visualizar datos.

Podemos establecer pues una genealogía y una temporización de todos estos lenguajes de marcado:



Así pues, el XML es una versión abreviada y simplificada del SGML optimizada para Internet. Si medimos la complejidad de las especificaciones por su número de páginas, las del SGML tiene unas 500 páginas mientras que las del XML ocupan sólo 80. El XML ofrece el 80% de las ventajas del SGML con sólo el 20% de las su complejidad.

5.7.b. Ejemplos sencillos de documentos XML

A continuación podemos observar un ejemplo de documento XML. Como ya hemos comentado, una de las características de este lenguaje es que su lectura es muy clara. Seguramente no te costará nada deducir qué tipo de información representa este archivo:

```
<?xml version = "1.0" standalone = "yes"?>
<diccionario>
  <entrada id='1'>
    <spa>casa</spa>
    <eng>house</eng>
  </entrada>
  <entrada id='2'>
    <cat>cotxe</cat>
    <spa>coche</spa>
    <eng>car</eng>
  </entrada>
</diccionario>
```

Tampoco te costará entender qué tipo de información contiene el siguiente ejemplo:

```
<?xml version = "1.0" standalone = "yes"?>
<agenda-telefonos>
  <contacto id='1'>
    <nombre>María Gil</nombre>
    <telefono>456783909</telefono>
  </contacto>
  <contacto id='2'>
    <nombre>Ernesto Villalba</nombre>
    <telefono>768436543</telefono>
  </contacto>
</agenda-telefonos>
```

Los dos ejemplos que hemos presentado son archivos XML bien formados, pero no se basan en ningún estándar para representar bases de datos terminológicas o agendas de teléfono. Como veremos más adelante, de XML se derivan una serie de estándares que definen cómo deben ser los documentos XML para una determinada aplicación.

5.7.c. Estructura de los documentos XML

Ya hemos comentado más arriba que un XML es un formato de texto. El texto puede tener dos funciones diferenciadas: o marcar o ser un dato. Las marcas de XML sirven para estructurar de una manera lógica el documento XML. Observa el siguiente ejemplo:

```
<?xml version = "1.0" standalone = "yes"?>
<libro id='143'>
  <titulo>Don Quijote de la Mancha</titulo>
  <autor>Miguel de Cervantes</autor>
</libro>
```

En este documento las marcas son: xml - version - standalone - libro - titulo - autor y los datos son: 1.0 - yes - Don Quijote de la Mancha - Miguel de Cervantes.

Elementos o tags

Todo lo que está entre los símbolos < y > se considera *elemento* (o *tag*) (excepto si está dentro de una sección CDATA). Los nombres de los elementos deben cumplir las siguientes normas:

- No puede empezar por cifras o caracteres de puntuación

- Puede contener letras, cifras y caracteres de puntuación (aunque es recomendable no utilizar el guión (-), ni los puntos (.) Ni los dos puntos (:))
- No puede empezar con xml (o XML, o Xml, etc)
- Los nombres no pueden contener espacios

Hay que recordar que los nombres son sensibles a las mayúsculas y minúsculas. Por tanto **<Nombre>** es diferente a **<nombre>**.

Un elemento puede estar vacío, es decir, que no contenga ningún dato. Entonces se puede cerrar con **</>**. Por ejemplo: **<entrada> </ entrada>** es equivalente a **<entrada />**.

Atributos

En el ejemplo anterior teníamos **<libro id = '143'>**. **id** es un atributo y el valor de este atributo es 143. Todo lo dicho para los nombres de los elementos también es válido para los atributos. Los valores de los atributos van entre comillas, tanto pueden ser comillas simples como dobles.

Comentarios

Los comentarios en XML se escriben igual que en HTML.

```
<! - Esto es un comentario en XML ->
```

Entidades

Son marcas que se reemplazan por caracteres cuando se analiza el documento. En XML sólo hay 5 entidades predefinidas:

| | |
|---------------------|-----------------------------|
| &amp; ; | & |
| &lt; ; | < |
| &gt; ; | > |
| &apos; ; | " (Comillas dobles) |
| &quot; ; | ' (Comillas simples) |

Bloques CDATA

Hasta ahora hemos dicho que todo lo que está entre los símbolos **<>** son marcas (o valores de atributo) y todo lo que está fuera son datos. Hay, sin embargo, una excepción las secciones o bloques CDATA. Lo que está dentro de una sección CDATA no es interpretado por el parser de XML. Los únicos caracteres no permitidos dentro de una sección CDATA son: **]]>** (ya que es la marca de cierre de las CDATA). La utilidad de estas secciones es hacer el documento más legible, veamos con un ejemplo:

Ejemplo: si tuviéramos que poner en un documento XML:

```
x <c2> y
```

deberíamos hacerlo con las entidades:

```
x &lt; c2 &gt; y
```

lo que resulta un poco difícil de leer. Podemos facilitar la lectura escribiendo:

```
<!CDATA[x < c2 > y]]>
```

5.7.d. Los documentos XML bien formados

Hay 6 reglas que hay que respetar para asegurarnos de que un documento XML está bien formado.

- Todo elemento que contenga datos debe tener un tag para abrirlo y un tag para cerrarlo: <autor> Miguel de Cervantes </ autor>
- Todo elemento que no contenga datos debe tener un tag único terminado con />:

- Debe haber un único elemento que contenga a todos los demás (raíz). En el ejemplo del diccionario este elemento era <libro>
- Los elementos deben estar anidados, no superpuestos:

```
<b> Este XML <i> no está </b> bien formado.</i>
<b> Este XML <i> si está </i> bien formado.</b>
```

- Los valores de los atributos van entre comillas (simples o dobles)
- Las únicas referencias a entidades permitidas son: & < > ' y "

Definición de los tipos de documentos

El XML es un lenguaje que permite intercambiar información. En el momento del intercambio surge la necesidad de validar los documentos, es decir, verificar que los documentos estén bien formados y que sean válidos. La diferencia entre bien formado y válido es la siguiente:

- Un documento XML bien formado es aquel que tiene una sintaxis correcta, es decir, que cumple las 6 reglas básicas que hemos expuesto más arriba.
- Un documento está bien formado si tiene la estructura que esperábamos, es decir si es el tipo de documento esperado.

Existen dos mecanismos para validar los documentos XML:

- Los DTD (*Document Type Definition*)
- XML Schema

Los DTD y los XML Schema especifican las reglas que definen la estructura de un documento XML. No entraremos en detalles, simplemente mostraremos un ejemplo de cada, correspondiente a nuestro XML de ejemplo:

DTD (Document Type Definition)

El DTD puede estar incluido dentro del propio XML:

```
<?xml version = "1.0"?>
<DOCTYPE note [
  <!ELEMENT libro (título, autor)>
  <!ELEMENT título (#PCDATA)>
  <!ELEMENT autor (#PCDATA)>
  <!ATTLIST libro
    src CDATA #REQUIRED
  >
```

```

]>
<libro id='143'>
  <titulo>Don Quijote de la Mancha</titulo>
  <autor>Miguel de Cervantes</autor>
</libro>

```

El DTD también puede estar en un archivo externo y poniendo una referencia al archivo XML:

El archivo XML:

```

<?xml version = "1.0"?>
<!DOCTYPE note SYSTEM "llibre.dtd">
<libro id ='143'>
  <titulo>Don Quijote de la Mancha</titulo>
  <autor>Miguel de Cervantes</autor>
</libro>

```

Y el DTD "llibre.dtd"

```

<?xml version = "1.0"?>
<DOCTYPE note [
  <!ELEMENTO libro (título, autor)>
  <!ELEMENTO titulo (#PCDATA)>
  <!ELEMENTO autor (#PCDATA)>
  <!ATTLIST libro
    src CDATA #REQUIRED
  >
]>

```

XML Schema

El XML Schema es a la vez un XML, que si os fijáis en el siguiente ejemplo, tiene un XML Schema que sirve para validarlo (está definido en la segunda línea).

```

<?xml version="1.0" encoding="utf-16"?>
<xsd:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
version="1.0" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="libro" type="llibreType" />
  <xsd:complexType name="llibreType">
    <xsd:sequence>
      <xsd:element name="titulo" type="xsd:string" />
      <xsd:element name="autor" type="xsd:string" />
    </xsd:sequence>
    <xsd:attribute name="id" type="xsd:int" />
  </xsd:complexType>
</xsd:schema>

```

El XML Schema se referencia dentro del propio XML

```

<?xml version = "1.0"?>
<libro id = '143' xmlns: XSI = "http://www.w3.org/2001/XMLSchema-instance"
XSI: noNamespaceSchemaLocation = "llibre.xsd">
  <titulo> Don Quijote de la Mancha </titulo>
  <autor> Miguel de Cervantes </autor>
</libro>

```

Como vemos, para definir el tipo de documento se puede usar tanto un DTD como un XML Schema. En general, sin embargo, se considera que el XML Schema es mucho más potente y es la opción recomendada.

5.7.e. Tecnologías asociadas: XSLT y XPath

En esta sección hablaremos de dos importantes tecnologías asociadas a XML: las XSLT o Transformaciones XSL (*Extensible Stylesheet Language Transformations*) y XPath (*XML Path Language*).

XSLT

Es un sofisticado lenguaje que permite transformar un XML en otro XML diferente, seleccionando qué información del XML original debe aparecer y de qué manera en el XML transformado. Una aplicación muy habitual del XSLT es transformar un XML en un XHTML para poderlo transmitir a través de Internet y visualizarlo en cualquier navegador. Veamos un ejemplo:

Disponemos del fichero to.xml, que contiene un glosario terminológico en el formato de Terminología Oberta del TermCat:

```
<?xml version="1.0" encoding="UTF-8"?>
<terminologiaoberta>
<autor>TERMCAT, Centre de Terminologia</autor>
<titol>TO Termes normalitzats (2007)</titol>
<fitxes>
<fitxa num="1">
<areatematica>Protecció civil: Policia</areatematica>
<denominacio llengua="ca" tipus="principal" jerarquia="terme pral." categoria="loc adj">a boca de canó</denominacio>
<denominacio llengua="es" tipus="equivalent" jerarquia="terme pral." categoria="">a bocaजारro </denominacio>
</fitxa>
<fitxa num="2">
<areatematica>Protecció civil: Policia</areatematica>
<denominacio llengua="ca" tipus="principal" jerarquia="terme pral." categoria="loc adj">a curta distància</denominacio>
<denominacio llengua="es" tipus="equivalent" jerarquia="terme pral." categoria="">a corta distancia</denominacio>
<denominacio llengua="en" tipus="equivalent" jerarquia="terme pral." categoria="">near </denominacio>
</fitxa>
....
<fitxa num="5880">
<areatematica>Esports: Esquí artístic i acrobàtic</areatematica>
<denominacio llengua="ca" tipus="principal" jerarquia="terme pral." categoria="m">zúdnic </denominacio>
<denominacio llengua="es" tipus="equivalent" jerarquia="terme pral." categoria="">zudnik </denominacio>
<denominacio llengua="fr" tipus="equivalent" jerarquia="terme pral." categoria="">zudnik </denominacio>
<denominacio llengua="en" tipus="equivalent" jerarquia="terme pral." categoria="">zudnik </denominacio>
</fitxa>
</fitxes>
</terminologiaoberta>
```

El siguiente archivo XSLT (stylesheetTO.xsl) transformará este XML en un HTML que mostrará una tabla con los términos en las diferentes lenguas:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
<html>
<body>
<h2>Terminologia</h2>
<table border="1">
<tr bgcolor="#9acd32">
<th align="left">ca</th>
<th align="left">en</th>
<th align="left">es</th>
<th align="right">Area temàtica</th>
</tr>
<xsl:for-each select="terminologiaoberta/fitxes/fitxa">
```

```

<tr>
  <td><xsl:value-of select="denominacio [@llengua='ca']"/></td>
  <td><xsl:value-of select="denominacio [@llengua='en']"/></td>
  <td><xsl:value-of select="denominacio [@llengua='es']"/></td>
  <td><xsl:value-of select="areatematica"/></td>
</tr>
</xsl:for-each>
</table>
</body>
</html>
</xsl:template>
</xsl:stylesheet>

```

Para aplicar esta hoja de estilo XSLT en el archivo XML podemos utilizar diversas herramientas, como por ejemplo xsltproc disponible en Linux bajo terminal:

```
xsltproc stylesheetTO.xsl to.xml > sortida.html
```

El fichero sortida.xhtml tendrá el siguiente aspecto:

```

<html>
<body>
<h2>Terminologia</h2>
<table border="1">
<tr bgcolor="#9acd32">
<th align="left">ca</th>
<th align="left">en</th>
<th align="left">es</th>
<th align="right">Area temàtica</th>
</tr>
<tr>
<td>a boca de canó</td>
<td></td>
<td>a bocajarro</td>
<td>Protecció civil: Policia</td>
</tr>
<tr>
<td>a curta distància</td>
<td>near</td>
<td>a corta distancia</td>
<td>Protecció civil: Policia</td>
</tr>
<tr>
<td>a frec de roba</td>
<td></td>
<td>a quemarropa</td>
<td>Protecció civil: Policia</td>
</tr>
...
<tr>
<td>zúdnic</td>
<td>zudnik</td>
<td>zudnik</td>
<td>Esports: Esquí artístic i acrobàtic</td>
</tr>
</table>
</body>
</html>

```

Si visualizamos este documento en un navegador de Internet veremos lo siguiente:

Terminologia

| ca | en | es | Area temàtica |
|-------------------------|----------------------------|--------------------------|---------------------------------------|
| a boca de canó | | a bocajarro | Protecció civil: Policia |
| a curta distància | near | a corta distancia | Protecció civil: Policia |
| a frec de roba | | a quemarropa | Protecció civil: Policia |
| a l'unison | in unison | | Esports: Patinatge artístic sobre gel |
| a llarga distància | distant | a larga distancia | Protecció civil: Policia |
| a peu de fàbrica | ex works | en fábrica | Comerç internacional |
| a portell | quincunx | a tresbolillo | Construcció |
| abadejo | pollack | abadejo | Peixos |
| abandonament | drop out | abandono | Farmacologia |
| abatre un obstacle | knock down an obstacle, to | derribar un obstáculo | Esports: Atletisme |
| abonament | ski-pass | abono | Esports d'hivern |
| aborrallament | pilling | pildeo | Indústria tèxtil: Teixits |
| abrusament | kindling | activación propagada | Psiquiatria |
| absortància | absortance | absortancia | Química física |
| accident en el trajecte | commuting accident | accidente en el trayecto | Assegurances |
| accidentogen -ògena | accident prone | accidentógeno | Transport per carretera |
| acció de cessació | | acción de cesación | Dret civil |
| acció de primera | blue chip | acción de primera clase | Borsa |
| acció del mos | action of the loin | acción del bocado | Esports: Hípica |
| acció negatòria | | acción negatoria | Dret civil |
| acer maràging | maraging steel | acero maraging | Indústria metal·lúrgica |

Para aprender más sobre XSLT se puede consultar el tutorial de W3Schools en: <http://www.w3schools.com/xsl/>

XPath

XPath es un lenguaje que permite seleccionar nodos y conjunto de nodos de un documento XML. Es un lenguaje sencillo, que recuerda a las instrucciones de línea de comandos de muchos sistemas operativos, pero que es muy potente y permite hacer consultas a un documento XML como si se tratara de una especie de base de datos.

Siguiendo con el ejemplo de XML de la base de datos terminológica del TermCat, podemos ver las siguientes expresiones y sus resultados. Para evaluar las expresiones utilizaremos xmllint, una aplicación en línea de comandos para Linux.

Para seleccionar todas las denominaciones podemos hacer:

```
xmllint -xpath '/terminologiaoberta/fitxes/fitxa/denominacio' to.xml
```

O bien:

```
xmllint -xpath '//denominacio' to.xml
```

El resultado que obtenemos son todas las denominaciones:

```
<denominacio llengua="ca" tipus="principal" jerarquia="terme pral."
categoria="
loc adj">a boca de canó</denominacio><denominacio llengua="es" tipus="equi
valent" jerarquia="terme pral." categoria="">a
bocajarro</denominacio><denomina
cio llengua="ca" tipus="principal" jerarquia="terme pral." categoria="loc
adj">
a curta distància</denominacio><denominacio llengua="es" tipus="equivalent
" jerarquia="terme pral." categoria="">a corta
distancia</denominacio><denomina
```



```

cio llengua="en" tipus="equivalent" jerarquia="terme pral."
categoria="">near</
denominacio><denominacio llengua="ca" tipus="principal" jerarquia="terme
pral."
  categoria="loc adj">a freq de roba</denominacio>.....

```

Si ahora queremos sólo las denominaciones en catalán podemos hacer:

```
xmllint -xpath '/terminologiaoberta/fitxes/fitxa/denominacio[@llengua="ca"]' to.xml
```

O bien:

```
xmllint -xpath '//denominacio[@llengua="ca"]' to.xml
```

El resultado que obtenemos en ambos casos son todas las denominaciones en catalán:

```

<denominacio llengua="ca" tipus="principal" jerarquia="terme pral."
categoria=""
loc adj">a boca de canó</denominacio><denominacio llengua="ca" tipus="prin
cipal" jerarquia="terme pral." categoria="loc adj">a curta distància</deno
minacio><denominacio llengua="ca" tipus="principal" jerarquia="terme pral."
cat
egoria="loc adj">a freq de roba</denominacio>...

```

Si ahora queremos la información sólo del segundo elemento podemos escribir:

```
xmllint -xpath '/terminologiaoberta/fitxes/fitxa[2]/denominacio[@llengua="ca"]' to.xml
```

Y obtenemos:

```

<denominacio llengua="ca" tipus="principal" jerarquia="terme pral."
categoria="loc adj">a curta distància</denominacio>

```

También podemos combinar dos expresiones, para por ejemplo obtener la denominación catalana y la inglesa, como por ejemplo:

```
xmllint -xpath '//fitxa[2]/denominacio[@llengua="ca"] |
//fitxa[2]/denominacio[@llengua="en"]' to.xml
```

Donde obtenemos:

```

<denominacio llengua="ca" tipus="principal" jerarquia="terme pral."
categoria="loc adj">a curta distància</denominacio><denominacio llengua="en"
tipus="equivalent" jerarquia="terme pral." categoria="">near</denominacio>

```

Otra muy buena opción de herramienta en línea de comandos, disponible para Linux y Windows, para evaluar XPath es xmlstarlet (<http://xmlstar.sourceforge.net/>). Si hacemos:

```
xmlstarlet sel -t -v '//denominacio' to.xml
```

obtenemos:

```

a bocajarro
a bocajarro
a curta distancia
a curta distancia

```

```
near
a quemarropa
a quema-ropa
a quemarropa
al unísono
```

Es decir, sin los tags de XML.

Puedes aprender más sobre las expresiones XPath al tutorial del W3Schools (<http://www.w3schools.com/xpath>).

5.7.f. Editores de XML

Los archivos XML son archivos de texto, y pueden editarse en cualquier editor de texto. No obstante, es recomendable utilizar editores específicos para XML. Estos editores específicos tienen varias ventajas:

- Colorean el texto diferenciando elementos, atributos, etc., lo que ayuda visualmente a escribir y leer los ficheros
- Avisan si dejamos de cerrar alguna etiqueta
- Completado automático de etiquetas
- Pueden verificar si el documento está bien formado y si es válido.
- Algunos de ellos permiten hacer también transformaciones XSLT
- Algunos de ellos permiten también evaluar expresiones XPath

Una buena opción, y de software libre, disponible tanto para Linux como para Windows, es el XML Copy Editor, que se puede descargar de <http://xml-copy-editor.sourceforge.net/>

5.7.g. Traducción de documentos XML

Los archivos XML son archivos de texto y por lo tanto no ofrecen demasiada dificultad en su traducción. La traducción se puede hacer directamente con cualquier editor de texto. Si utilizamos un editor específico para XML nos podrá avisar si en algún momento modificamos algún aspecto de la estructura del XML.

Los archivos XML también se pueden traducir fácilmente con cualquier herramienta de traducción asistida. Imaginemos que tenemos el siguiente archivo XML para traducir (mostramos sólo la información correspondiente a uno de los libros y además de forma abreviada, pero imaginemos que hay cientos de estos):

```
<bd_libros>
<libro id="1">
  <autor>John Steinbeck</autor>
  <titulo>Las uvas de la ira</titulo>
  <traductor>Mercedes López arrabio</traductor>
  <editorial> Ediciones 62 </editorial>
  <coleccion>Las mejores obras de la literatura universal siglo
  XX</coleccion>
  <numero>83</numero>
  <ano_publicacion> 1993 </ano_publicacion>
  <biografia_autor>John Steinbeck (1902-1968), escritor norteamericano,
  nacido en California. ...</biografia_autor>
```

```

    <resumen>Cuando se publicó Las uvas de la ira en 1939, conmueve la
    opinión norteamericana, que se estaba apenas recuperando de la Gran
    Depresión. ....</resumen>
</libro>
...
</bd_libros>

```

Si creamos un archivo XLIFF sin crear un filtro específico (en esta prueba hemos utilizado la herramienta Rainbow de Okapi) obtenemos lo siguiente:

```

<?xml version = "1.0" encoding = "UTF-8"?>
<xliff version = "1.2" xmlns = "urn: oasis: names: tc: XLIFF: documento: 1.2"
xmlns: okp = "okapi-framework: XLIFF-extensiones" xmlns: its = "http:
//www.w3. org / 2005/11 / its "xmlns: itsxlf = " http://www.w3.org/ns/its-
xliff/ "its: version =" 2.0 ">
<file original = "bdlibros.xml" source-language = "es" target-language = "ca"
datatype = "xml">
<body>
<trans-unit id = "1">
<source xml: lang = "es"> John Steinbeck </source>
<target xml: lang = "ca"> John Steinbeck </target>
</trans-unit>
<trans-unit id = "2">
<source xml: lang = "es"> Las uvas de la ira </source>
<target xml: lang = "ca"> Las uvas de la ira </target>
</trans-unit>
<trans-unit id = "3">
<source xml: lang = "es"> Mercedes López arrabio </source>
<target xml: lang = "ca"> Mercedes López arrabio </target>
</trans-unit>
<trans-unit id = "4">
<source xml: lang = "es"> Ediciones 62 </source>
<target xml: lang = "ca"> Ediciones 62 </target>
</trans-unit>
<trans-unit id = "5">
<source xml: lang = "es"> Las mejores obras de la literatura universal siglo
XX </source>
<target xml: lang = "ca"> Las mejores obras de la literatura universal siglo
XX </target>
</trans-unit>
<trans-unit id = "6">
<source xml: lang = "es"> 83 </source>
<target xml: lang = "ca"> 83 </target>
</trans-unit>
<trans-unit id = "7">
<source xml: lang = "es"> 1993 </source>
<target xml: lang = "ca"> 1993 </target>
</trans-unit>
<trans-unit id = "8">
<source xml: lang = "es"> John Steinbeck (1902-1968), escritor
norteamericano, nacido en California. <source>
<target xml: lang = "ca"> John Steinbeck (1902-1968), escritor
norteamericano, nacido en California. </target>
</trans-unit>
<trans-unit id = "9">
<source xml: lang = "es"> Cuando se publicó Las uvas de la ira en 1939,
conmueve la opinión norteamericana, que se estaba apenas recuperando de la
Gran Depresión. </source>

```

```
<target xml: lang = "ca"> Cuando se publicó Las uvas de la ira en 1939,
conmueve la opinión norteamericana, que se estaba apenas recuperando de la
Gran Depresión.</target>
</trans-unit>
</body>
</file>
</xliff>
```

Para poder traducir de manera eficiente este archivo tendremos que crear un filtro específico para que el programa seleccione sólo la información que es traducible, y evitar tener que traducir por ejemplo, el nombre del autor, del traductor, la editorial, la colección y el año de publicación. Quizás para unos pocos libros no tiene importancia, pero si el archivo XML tiene una gran cantidad de entradas puede ser interesante eliminar esta información del archivo a traducir. Así pues, sólo queremos traducir el título, la biografía_autor y el resumen.

Para hacer esto podemos crear un filtro específico para este archivo XML. Rainbow utiliza la recomendación ITS (*Internationalization Tag Set*) para crear los filtros. Se puede consultar los detalles en: http://www.opentag.com/okapi/wiki/index.php?title=XML_Filter. El filtro resultante sería:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?><its:rules
xmlns:its="http://www.w3.org/2005/11/its" xmlns:itsx="http://www.w3.org/2008/12/its-extensions"
xmlns:okp="okapi-framework:xmlfilter-options" xmlns:xlink="http://www.w3.org/1999/xlink"
version="1.0">
<its:translateRule selector="//autor" translate="no"/>
<its:translateRule selector="//traductor" translate="no"/>
<its:translateRule selector="//editorial" translate="no"/>
<its:translateRule selector="//colleccio" translate="no"/>
<its:translateRule selector="//numero" translate="no"/>
<its:translateRule selector="//any_publicacio" translate="no"/>
</its:rules>
```

Que si lo aplicamos a la creación del archivo XLIFF, nos resulta:

```
<?xml version = "1.0" encoding = "UTF-8"?>
<xliff version = "1.2" xmlns = "urn: oasis: names: tc: XLIFF: documento: 1.2"
xmlns: okp = "okapi-framework: XLIFF-extensiones" xmlns: its = "http:
//www.w3. org / 2005/11 / its "xmlns: itsxlf = " http://www.w3.org/ns/its-
xliff/ "its: version = " 2.0 ">
<file original = "bdlibros.xml" source-language = "es" target-language = "ca"
datatype = "xml">
<body>
<trans-unit id = "1">
<source xml: lang = "es"> Las uvas de la ira </source>
<target xml: lang = "ca"> Las uvas de la ira </target>
</trans-unit>
<trans-unit id = "2">
<source xml: lang = "es"> John Steinbeck (1902-1968), escritor
norteamericano, nacido en California. <source>
<target xml: lang = "ca"> John Steinbeck (1902-1968), escritor
norteamericano, nacido en California. </target>
</trans-unit>
<trans-unit id = "3">
<source xml: lang = "es"> Cuando se publicó Las uvas de la ira en 1939,
conmueve la opinión norteamericana, que se estaba apenas recuperando de la
Gran Depresión. </source>
<target xml: lang = "ca"> Cuando se publicó Las uvas de la ira en 1939,
conmueve la opinión norteamericana, que se estaba apenas recuperando de la
Gran Depresión.</target>
</trans-unit>
</body>
```

```
</file>  
</xliff>
```

Que contiene, efectivamente la información que queremos traducir. Cada herramienta de traducción asistida puede tener un mecanismo diferente para la creación de filtros XML, pero en todo caso siempre se trata de especificar qué información del XML es traducible, o bien, qué información no lo es.

5.8. Los formatos XML utilizados en el mundo de la traducción

En capítulos anteriores ya hemos visto algunos de los formatos de intercambio basados en XML que se utilizan en el mundo de la traducción. En esta sección revisaremos estos formatos y veremos también unas métricas especiales (GMX), que a pesar de no ser un formato XML, tienen relación con los nuevos estándares que están apareciendo en los últimos años.

5.8.a. Intercambio de memorias de traducción: TMX

El TMX (*Translation Memory eXchange*) es un formato XML estándar para el intercambio de memorias de traducción (el siguiente ejemplo no corresponde a un documento completo, únicamente a una entrada).

```
<tu tuid="1" datatype="Text" srclang="ca">
  <prop type="x-Client">001</prop>
  <prop type="x-Domain">0049</prop>
  <prop type="x-Project">2053797</prop>
  <prop type="FileID">1</prop>
  <prop type="RowID">0000009</prop>
  <tuv xml:lang="ca" creationdate="20030601T08:21:33Z"
    creationid="Antoni">
    <prop type="IsSource">True</prop>
    <seg>Tema 4.- </seg>
  </tuv>
  <tuv xml:lang="es" creationdate="20030601T08:21:33Z"
    creationid="Antoni">
    <prop type="IsSource">False</prop>
    <seg>Tema 4.-</seg>
  </tuv>
  <tuv xml:lang="en-gb" creationdate="18991229T23:00:00Z"
    creationid="Antoni">
    <prop type="IsSource">False</prop>
    <seg>Unit 4.-</seg>
  </tuv>
</tu>
```

Esta memoria de traducción es multilingüe: catalán, castellano e inglés. La lengua original del segmento que presentamos es el catalán. Los datos que se guardan son cliente, especialidad, id. del proyecto de donde proviene y del segmento dentro del proyecto, fecha de creación, usuario que lo ha creado (en este caso Antoni) y el segmento en cada una de las lenguas.

5.8.b. Intercambio de bases de datos terminológicas: TBX

El TBX (*TermBase eXchange*) es un formato estándar para el intercambio de bases de datos terminológicas basado en XML. A continuación podemos observar un ejemplo sencillo:

```
<?xml version="1.0" ?>
<martif type="TBX" xml:lang="en">
<text>
<body>
  <termEntry id="1">
    <descrip type="subjectField">Linguistics</descrip>
    <langSet xml:lang="ru">
      <tig>
        <term>Компьютерная лингвистика</term>
      </tig>
    </langSet>
    <langSet xml:lang="en">
      <tig>
        <term>Computational linguistics</term>
      </tig>
    </langSet>
  </termEntry>
</body>
</text>
</martif>
```

```

</tig>
</langSet>
</termEntry>

```

5.8.c. Intercambio de proyectos de traducción: XLIFF

El XLIFF (*XML Localization Interchange File Format*) es un formato basado en XML para el intercambio de proyectos de traducción y localización. Con este formato es posible crear un proyecto con una herramienta de traducción asistida (A) y traducirlo con otra distinta (B), y una vez traducido, exportar el proyecto (es decir crear los archivos traducidos en su formato original) con la herramienta A.

```

<?xml version="1.0" encoding="UTF-8" ?>
<xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2">
<file datatype="x-test" original="manual"
source-language="EN-US" target-language="CA-ES">
<body>
  <trans-unit id="1">
    <source xml:lang="EN-US">Untranslated text.</source>
  </trans-unit>
  <trans-unit id="2">
    <source xml:lang="EN-US">Translated but un-approved text.</source>
    <target xml:lang="CA-ES">Text traduït però que encara no està aprovat.</target>
  </trans-unit>
  <trans-unit id="3" approved="yes">
    <source xml:lang="EN-US">Translated and approved text.</source>
    <target xml:lang="CA-ES">Text traduït i aprovat.</target>
  </trans-unit>
  <trans-unit id="4">
    <source xml:lang="EN-US">Some other text.</source>
    <alt-trans>
      <source xml:lang="EN-US">Other text.</source>
      <target xml:lang="CA-ES">Un altre text.</target>
    </alt-trans>
  </trans-unit>
</body>
</file>
</xliff>

```

5.8d. Intercambio de reglas de segmentación: SRX

SRX (*Segmentation Rule eXchange*) es un formato estándar basado en XML para el intercambio de reglas de segmentación, es decir, las reglas que se utilizan para dividir el texto a traducir en segmentos que se presentan de uno a uno al traductor. Si queremos aprovechar una determinada memoria de traducción es interesante que las reglas de segmentación que utilice nuestro programa de traducción asistida sean iguales a las reglas de segmentación utilizadas en la creación de la memoria de traducción. Si no es así, no es demasiado grave pero es posible que perdamos alguna coincidencia interesante por diferencias en la segmentación. Por este motivo se ha creado este formato, de modo que cuando compartimos nuestras memorias de traducción podamos también compartir las reglas de segmentación que utilizamos. La situación sería la siguiente. Un colega nuestro trabaja con la herramienta de traducción asistida A (que utiliza unas reglas de segmentación determinadas) y tiene una gran memoria de traducción que puede compartir con nosotros. Ahora nosotros queremos crear un proyecto de traducción con una herramienta de traducción asistida B (que utiliza unas reglas de segmentación que pueden ser diferentes que las de la herramienta A) pero queremos utilizar la gran memoria de traducción de nuestro colega. Para maximizar el número de coincidencias, será interesante pedir también el archivo de reglas de segmentación en formato SRX para poder crear el proyecto utilizando estas reglas.

Un archivo SRX tiene el siguiente aspecto:

```
<?xml version="1.0" encoding="UTF-8"?>
<srx      xmlns="http://www.lisa.org/srx20"      xmlns:okpsrx="http://okapi.sf.net/srx-extensions"
version="2.0">
<body>
<languageules>
<languageule languageulename="default">
<rule break="no">
<beforebreak>\b(pp|e\.\s*g|i\.\s*e|no|[Vv]ol|[Rr]col|maj|Lt|[Ff]ig|[Ff]igs|[Vv]iz|[Vv]ols|
[Aa]pprox|[Ii]ncl|Pres|Prof|[Dd]ept|min|max|[Gg]ovt|c\.\s*f|vs)\.</beforebreak>
<afterbreak>\s^\p{Lu}</afterbreak>
</rule>
<rule break="no">
<beforebreak>\b(St|Gen|Hon|Dr|Mr|Ms|Mrs|Col|Maj|Brig|Sgt|Capt|Cmnd|Sen|Rev|Rep|
Revd)\.</beforebreak>
<afterbreak>\s\p{Lu}</afterbreak>
</rule>
<rule break="no">
<beforebreak>([A-Z]\.){2,}</beforebreak>
<afterbreak>\s\p{Lu}</afterbreak>
</rule>
<rule break="yes">
<beforebreak>\w+[\p{Pe}\p{Po}]*[\.\?!]+[\p{Pe}\p{Po}]*</beforebreak>
<afterbreak>\s</afterbreak>
</rule>
</languageule>
</languageules>
<maprules>
<languageap languagepattern=".*" languageulename="default"></languageap>
</maprules>
</body>
</srx>
```

5.8.e. Métricas GILT: GMX

El GMX (*Global information management Metrics eXchange*) es una colección de estándares, algunos todavía en fase de propuesta, orientados principalmente a las necesidades de la industria de la traducción y que tienen que ver con la medida de los aspectos cuantitativos de un documento, especialmente aquellos que tienen una relevancia especial para el proceso de traducción (por ejemplo, recuento de palabras, complejidad, etc). No se trata, pues, de un formato, sino de una serie de procedimientos para poder medir de una manera unificada algunos aspectos de los documentos.

Podemos distinguir tres:

- GMX-V (que tiene que ver con el volumen) y establece una manera verificable de calcular los recuentos de palabras. Podemos ver el siguiente ejemplo: si tenemos el siguiente fragmento:

```
<source>In this <g id="g1">exa<x id="x1"/>mple</g> the in-line codes do not form part of
the word or character counts but are counted separately.</source>
```

se contaría como:

```
<source>In this example the in-line codes do not form part of the word or character counts
but are counted separately.</source>
```

y las medidas GMX- serían las siguientes:

```
words: 20, characters: 91, inline elements: 3, punctuation characters: 1, white space
characters: 19
```

Las especificaciones del GMX-V se pueden encontrar en:

<http://www.gala-global.org/oscarStandards/gmx-v/gmx-v.html>

- GMX-Q (que tiene que ver con la calidad) y que representa el nivel de calidad requerido para una determinada tarea de traducción. En el momento de escribir este capítulo aún no estaban disponibles las especificaciones.
- GMX-C (que tiene que ver con la complejidad) y que tendrá en cuenta el documento original, su formato. En el momento de escribir este capítulo aún no estaban disponibles las especificaciones.

6. Conclusiones

En este capítulo hemos visto los conceptos básicos necesarios para evitar problemas con las codificaciones de caracteres y los formatos de archivo. Una vez asimilados estos conceptos seremos capaces de traducir archivos en una gran cantidad de formatos y asegurándonos de que una vez traducidos los documentos se podrán abrir y visualizar correctamente.

Hemos dado una importancia especial al formato XML, ya que este formato se está extendiendo cada día más en todo tipo de aplicaciones. Con los conceptos adquiridos, seremos capaces de traducir archivos XML con nuestra herramienta de traducción asistida preferida.

Para ampliar conocimientos

Tutoriales de W3Schools

En este capítulo he hecho referencia a un tutorial de XML de W3Schools (<http://www.w3schools.com/>). En esta web se ofrecen tutoriales gratuitos sobre diversas tecnologías relacionadas con Internet. Vale la pena darle un vistazo y seguir los tutoriales de los temas que te interesen. Estos tutoriales están muy bien desarrollados, son claros y directos, y te permiten tener una buena idea de estas tecnologías en muy poco tiempo.

Detección automática de lengua

Hemos comentado que la detección automática de la codificación de caracteres se lleva a cabo de una manera heurística a partir de estadísticas de trigramas de caracteres. Una técnica similar se puede utilizar para determinar la lengua en la que está escrito un documento. De esta manera se pueden crear detectores automáticos de lengua, que a partir de fragmentos de un texto son capaces de determinar con bastante precisión la lengua en la que está escrito.

Hay varias webs que permiten detectar la lengua, entre las que podemos destacar:

- Language Identifier Xerox (<https://open.xerox.com/Services/LanguageIdentifier>)
- TextCat (<http://odur.let.rug.nl/~vannoord/TextCat/>). Esta permite descargar el programa y los modelos de lengua.

Si alguna vez os llega un texto y no estáis seguros de en qué lengua está escrito podréis utilizar una de estas herramientas.

Tablas de Unicode

En <http://www.unicode.org/charts/> se pueden consultar todas las tablas de caracteres. Mostramos ahora algunas de las posiciones del Unicode, correspondientes al silabario japonés hiragana ya algunos de carácter CJK (chinos, japoneses y coreanos unificados):

12353 -> □ -> hiragana LETTER SMALL A
12354 -> □ -> hiragana LETTER A
12355 -> □ -> hiragana LETTER SMALL Y
12356 -> □ -> hiragana LETTER Y
12357 -> □ -> hiragana LETTER SMALL U
12358 -> □ -> hiragana LETTER U
12359 -> □ -> hiragana LETTER SMALL E
12360 -> □ -> hiragana LETTER E
12361 -> □ -> hiragana LETTER SMALL O
12362 -> □ -> hiragana LETTER O
12363 -> □ -> hiragana LETTER KA
12364 -> □ -> hiragana LETTER GA
12365 -> □ -> hiragana LETTER KI
12366 -> □ -> hiragana LETTER GI
12367 -> □ -> hiragana LETTER KU
12368 -> □ -> hiragana LETTER GU
12369 -> □ -> hiragana LETTER KE
12370 -> □ -> hiragana LETTER GE
12371 -> □ -> hiragana LETTER KO
12372 -> □ -> hiragana LETTER GO
12373 -> □ -> hiragana LETTER SA
12374 -> □ -> hiragana LETTER ZA
12375 -> □ -> hiragana LETTER SI
12376 -> □ -> hiragana LETTER ZI
12377 -> □ -> hiragana LETTER SU
12378 -> □ -> hiragana LETTER ZU
12379 -> □ -> hiragana LETTER SE
12380 -> □ -> hiragana LETTER ZE
12381 -> □ -> hiragana LETTER SO
12382 -> □ -> hiragana LETTER ZO
12383 -> □ -> hiragana LETTER TA
12384 -> □ -> hiragana LETTER DA
12385 -> □ -> hiragana LETTER TI
12386 -> □ -> hiragana LETTER DI
12387 -> □ -> hiragana LETTER SMALL TI
12388 -> □ -> hiragana LETTER TI
12389 -> □ -> hiragana LETTER DU
12390 -> □ -> hiragana LETTER TE
12391 -> □ -> hiragana LETTER DE
12392 -> □ -> hiragana LETTER TO
12393 -> □ -> hiragana LETTER DO
12394 -> □ -> hiragana LETTER NA
12395 -> □ -> hiragana LETTER NI
12396 -> □ -> hiragana LETTER NU
12397 -> □ -> hiragana LETTER NE
12398 -> □ -> hiragana LETTER NO
12399 -> □ -> hiragana LETTER HA

12400 -> -> hiragana LETTER BA
12401 -> -> hiragana LETTER PA
12402 -> -> hiragana LETTER HI
12403 -> -> hiragana LETTER BI
12404 -> -> hiragana LETTER PI
12405 -> -> hiragana LETTER HU
12406 -> -> hiragana LETTER BU
12407 -> -> hiragana LETTER PU
12408 -> -> hiragana LETTER HE
12409 -> -> hiragana LETTER BE
12410 -> -> hiragana LETTER PE
12411 -> -> hiragana LETTER HO
12412 -> -> hiragana LETTER BO
12413 -> -> hiragana LETTER PO
12414 -> -> hiragana LETTER MA
12415 -> -> hiragana LETTER MI
12416 -> -> hiragana LETTER MU
12417 -> -> hiragana LETTER ME
12418 -> -> hiragana LETTER MO
12419 -> -> hiragana LETTER SMALL YA
12420 -> -> hiragana LETTER YA
12421 -> -> hiragana LETTER SMALL YU
12422 -> -> hiragana LETTER YU
12423 -> -> hiragana LETTER SMALL YO
12424 -> -> hiragana LETTER YO
12425 -> -> hiragana LETTER RA
12426 -> -> hiragana LETTER RI
12427 -> -> hiragana LETTER RU
12428 -> -> hiragana LETTER RE
12429 -> -> hiragana LETTER RO
12430 -> -> hiragana LETTER SMALL WA
12431 -> -> hiragana LETTER WA
12432 -> -> hiragana LETTER WI
12433 -> -> hiragana LETTER WE
12434 -> -> hiragana LETTER WO
12435 -> -> hiragana LETTER N
12436 -> -> hiragana LETTER VU
12437 -> -> hiragana LETTER SMALL KA
12438 -> -> hiragana LETTER SMALL KE
....
23473 -> -> CJK UNIFIED IDEOGRAPH-5BB1
23474 -> -> CJK UNIFIED IDEOGRAPH-5BB2
23475 -> -> CJK UNIFIED IDEOGRAPH-5BB3
23.476 -> -> CJK UNIFIED IDEOGRAPH-5BB4
23477 -> -> CJK UNIFIED IDEOGRAPH-5BB5
23478 -> -> CJK UNIFIED IDEOGRAPH-5BB6
23479 -> -> CJK UNIFIED IDEOGRAPH-5BB7
23480 -> -> CJK UNIFIED IDEOGRAPH-5BB8
23481 -> -> CJK UNIFIED IDEOGRAPH-5BB9
23482 -> -> CJK UNIFIED IDEOGRAPH-5BBA

Problemas de visualización de documentos relacionados con las fuentes

El Unicode es capaz de representar la mayoría de caracteres existentes, pero eso no quiere decir que los visualizamos correctamente. Para poder visualizarlos debemos tener una fuente que tenga el glifo (la representación del carácter). Fijate en este ejemplo, de un texto en glagolítico:

□□□□□□□□□□□□□□□□

Desde el punto de vista informático, no se ha perdido ninguna información (ya que en esa posición están los bytes que debe haber), pero el editor no es capaz de visualizar el contenido. Si seleccionamos una fuente que contenga los caracteres a representar, entonces podremos ver el texto correspondiente:

□□□□□□□□ □□□□□ □

También, si exploramos algunas de las zonas del Unicode, como en el siguiente ejemplo, veremos que no tenemos los caracteres correspondientes.

18558 -->  -->CJK UNIFIED IDEOGRAPH-487E

que corresponde a:



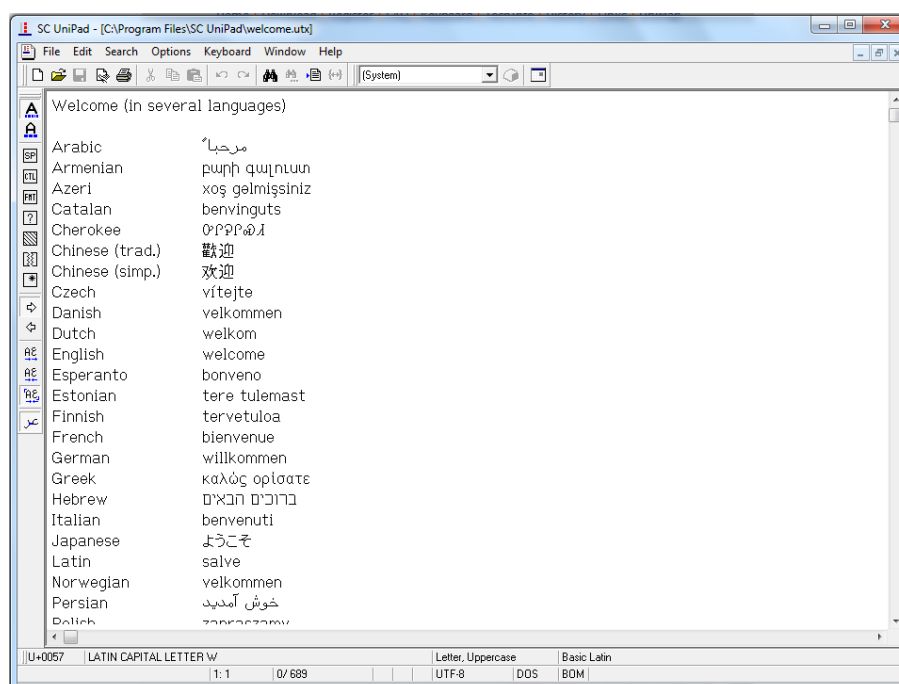
Así que recordemos que el hecho de ver en nuestro editor un cuadradito en lugar del carácter adecuado supone que la fuente que estamos usando no tiene la imagen del carácter correspondiente. Podremos seleccionar otra fuente e incluso buscar nuevas fuentes para instalar en nuestro sistema.

SC Unipad: Editor de Unicode

Los archivos Unicode son archivos de texto que se pueden abrir en cualquier editor de textos. Hay sin embargo, unos editores especiales para Unicode que simplemente ofrecen las siguientes funcionalidades:

- Trabajan con una fuente completa de todo el Unicode, por lo que pueden visualizar la totalidad de los caracteres representados por el Unicode.
- Ofrecen teclados virtuales para poder entrar el texto fácilmente en diversos alfabetos
- Permiten seleccionar cualquier carácter del repertorio Unicode.

SCP-Unipad es un buen ejemplo de este tipo de editores. Está disponible para Windows, es gratuito aunque no de software libre y se puede descargar de <http://www.unipad.org/>.



Bibliografía

Sayton, Bob (2007) *DocBook XSL: The Complete Guide* Sagehill Enterprises. Se puede acceder al contenido de este libro a <http://sagehill.net/docbookxsl/index.html>

Anexo I. Entidades de html

Caracteres ASCII

| | Entity Name | Entity Number |
|----|--------------------|----------------------|
| | | |
| ! | | ! |
| " | | " |
| # | | # |
| \$ | | $ |
| % | | % |
| & | & | & |
| ' | | ' |
| (| | (|
|) | |) |
| * | | * |
| + | | + |
| , | | , |
| - | | - |
| . | | . |
| / | | / |
| 0 | | 0 |
| 1 | | 1 |
| 2 | | 2 |
| 3 | | 3 |
| 4 | | 4 |
| 5 | | 5 |
| 6 | | 6 |
| 7 | | 7 |
| 8 | | 8 |
| 9 | | 9 |
| : | | : |
| ; | | ; |
| < | < | < |
| = | | = |
| > | > | > |
| ? | | ? |
| @ | | @ |
| A | | A |
| B | | B |

| Entity Name | Entity Number |
|--------------------|----------------------|
| C | C |
| D | D |
| E | E |
| F | F |
| G | G |
| H | H |
| I | I |
| J | J |
| K | K |
| L | L |
| M | M |
| N | N |
| O | O |
| P | P |
| Q | Q |
| R | R |
| S | S |
| T | T |
| U | U |
| V | V |
| W | W |
| X | X |
| Y | Y |
| Z | Z |
| [| [|
| \ | \ |
|] |] |
| ^ | ^ |
| _ | _ |
| ` | ` |
| a | a |
| b | b |
| c | c |
| d | d |
| e | e |
| f | f |
| g | g |
| h | h |
| i | i |
| j | j |
| k | k |
| l | l |
| m | m |
| n | n |
| o | o |
| p | p |
| q | q |
| r | r |
| s | s |

| Entity Name | Entity Number |
|-------------|---------------|
| t | t |
| u | u |
| v | v |
| w | w |
| x | x |
| y | y |
| z | z |
| { | { |
| | | |
| } | } |
| ~ | ~ |

Caracteres ISO-8859-1

| Entity Name | Entity Number |
|-------------|-----------------|
| À | À À |
| Á | Á Á |
| Â | Â Â |
| Ã | Ã Ã |
| Ä | Ä Ä |
| Å | Å Å |
| Æ | Æ Æ |
| Ç | Ç Ç |
| È | È È |
| É | É É |
| Ê | Ê Ê |
| Ë | Ë Ë |
| Ì | Ì Ì |
| Í | Í Í |
| Î | Î Î |
| Ï | Ï Ï |
| Ð | Ð Ð |
| Ñ | Ñ Ñ |
| Ò | Ò Ò |
| Ó | Ó Ó |
| Ô | Ô Ô |
| Õ | Õ Õ |
| Ö | Ö Ö |
| Ø | Ø Ø |
| Ù | Ù Ù |
| Ú | Ú Ú |
| Û | Û Û |
| Ü | Ü Ü |
| Ý | Ý Ý |
| Þ | Þ Þ |
| ß | ß ß |
| à | à à |
| á | á á |

| | Entity Name | Entity Number |
|---|--------------------|----------------------|
| â | â | â |
| ã | ã | ã |
| ä | ä | ä |
| å | å | å |
| æ | æ | æ |
| ç | ç | ç |
| è | è | è |
| é | é | é |
| ê | ê | ê |
| ë | ë | ë |
| ì | ì | ì |
| í | í | í |
| î | î | î |
| ï | ï | ï |
| ð | ð | ð |
| ñ | ñ | ñ |
| ò | ò | ò |
| ó | ó | ó |
| ô | ô | ô |
| õ | õ | õ |
| ö | ö | ö |
| ø | ø | ø |
| ù | ù | ù |
| ú | ú | ú |
| û | û | û |
| ü | ü | ü |
| ý | ý | ý |
| þ | þ | þ |
| ÿ | ÿ | ÿ |

Símbolos ISO-8859-1

| | Entity Name | Entity Number |
|---|--------------------|----------------------|
| | | |
| ! | ¡ | ¡ |
| ¢ | ¢ | ¢ |
| £ | £ | £ |
| ¤ | ¤ | ¤ |
| ¥ | ¥ | ¥ |
| ¦ | ¦ | ¦ |
| § | § | § |
| ¨ | ¨ | ¨ |
| © | © | © |
| ª | ª | ª |
| « | « | « |
| ¬ | ¬ | ¬ |
|  | ­ | ­ |
| ® | ® | ® |
|  | ¯ | ¯ |

| | Entity Name | Entity Number |
|---|--------------------|----------------------|
| ° | ° | ° |
| ± | ± | ± |
| ² | ² | ² |
| ³ | ³ | ³ |
| ´ | ´ | ´ |
| μ | µ | µ |
| ¶ | ¶ | ¶ |
| ¸ | ¸ | ¸ |
| ¹ | ¹ | ¹ |
| º | º | º |
| » | » | » |
| ¼ | ¼ | ¼ |
| ½ | ½ | ½ |
| ¾ | ¾ | ¾ |
| ¿ | ¿ | ¿ |
| × | × | × |
| ÷ | ÷ | ÷ |

Símbolos matemàtics

| | Entity Name | Entity Number |
|---|--------------------|----------------------|
| ∀ | ∀ | ∀ |
| ∂ | ∂ | ∂ |
| ∃ | ∃ | ∃ |
| ∅ | ∅ | ∅ |
| ∇ | ∇ | ∇ |
| ∈ | ∈ | ∈ |
| ∉ | ∉ | ∉ |
| ∋ | ∋ | ∋ |
| ∏ | ∏ | ∏ |
| ∑ | ∑ | ∑ |
| − | − | − |
| * | ∗ | ∗ |
| √ | √ | √ |
| ∝ | ∝ | ∝ |
| ∞ | ∞ | ∞ |
| ∠ | ∠ | ∠ |
| ∧ | ∧ | ∧ |
| ∨ | ∨ | ∨ |
| ∩ | ∩ | ∩ |
| ∪ | ∪ | ∪ |
| ∫ | ∫ | ∫ |
| ∴ | ∴ | ∴ |
| ~ | ∼ | ∼ |
| ≡ | ≅ | ≅ |
| ≈ | ≈ | ≈ |
| ≠ | ≠ | ≠ |
| ≡ | ≡ | ≡ |
| ≤ | ≤ | ≤ |

| | Entity Name | Entity Number |
|---|--------------------|----------------------|
| ≥ | ≥ | ≥ |
| ⊂ | ⊂ | ⊂ |
| ⊃ | ⊃ | ⊃ |
| ⊄ | ⊅ | ⊄ |
| ⊆ | ⊆ | ⊆ |
| ⊇ | ⊇ | ⊇ |
| ⊕ | ⊕ | ⊕ |
| ⊗ | ⊗ | ⊗ |
| ⊥ | ⊥ | ⊥ |
| · | ⋅ | ⋅ |

Letras griegas

| Letter | Entity Name | Entity Number |
|---------------|--------------------|----------------------|
| A | Α | Α |
| B | Β | Β |
| Γ | Γ | Γ |
| Δ | Δ | Δ |
| E | Ε | Ε |
| Z | Ζ | Ζ |
| H | Η | Η |
| Θ | Θ | Θ |
| I | ι | Ι |
| K | Κ | Κ |
| Λ | Λ | Λ |
| M | Μ | Μ |
| N | Ν | Ν |
| Ξ | Ξ | Ξ |
| O | Ο | Ο |
| Π | Π | Π |
| P | Ρ | Ρ |
| Σ | Σ | Σ |
| T | Τ | Τ |
| Υ | Υ | Υ |
| Φ | Φ | Φ |
| X | Χ | Χ |
| Ψ | Ψ | Ψ |
| Ω | Ω | Ω |
| α | α | α |
| β | β | β |
| γ | γ | γ |
| δ | δ | δ |
| ε | ε | ε |
| ζ | ζ | ζ |
| η | η | η |
| θ | θ | θ |
| ι | ι | ι |
| κ | κ | κ |
| λ | λ | λ |

Letter Entity Name Entity Number

| | | |
|---|------------|--------|
| μ | μ | μ |
| ν | ν | ν |
| ξ | ξ | ξ |
| ο | ο | ο |
| π | π | π |
| ρ | ρ | ρ |
| ς | ς | ς |
| σ | σ | σ |
| σ | σ | σ |
| τ | τ | τ |
| υ | υ | υ |
| φ | φ | φ |
| χ | χ | χ |
| ψ | ψ | ψ |
| ω | ω | ω |
| θ | ϑ | ϑ |
| Υ | ϒ | ϒ |
| Ϝ | ϖ | ϖ |

Otros símbolos
Symbol Entity Name Entity Number

| | | |
|---|----------|---------|
| Œ | Œ | Œ |
| œ | œ | œ |
| Š | Š | Š |
| š | š | š |
| Ÿ | Ÿ | Ÿ |
| f | ƒ | ƒ |
| ^ | ˆ | ˆ |
| ~ | ˜ | ˜ |
| |   |   |
| |   |   |
| |   |   |
| | ‌ | ‌ |
| | ‍ | ‍ |
| | &lrn; | ‎ |
| | ‏ | ‏ |
| - | – | – |
| — | — | — |
| ‘ | ‘ | ‘ |
| ’ | ’ | ’ |
| , | ‚ | ‚ |
| “ | “ | “ |
| ” | ” | ” |
| „ | „ | „ |
| † | † | † |
| ‡ | ‡ | ‡ |
| • | • | • |

| Symbol | Entity Name | Entity Number |
|--------|-------------|---------------|
| ... | … | … |
| ‰ | ‰ | ‰ |
| ´ | ′ | ′ |
| ” | ″ | ″ |
| < | ‹ | ‹ |
| > | › | ‹ |
| – | ‾ | ‾ |
| € | € | € |
| ™ | ™ | ™ |
| ← | ← | ← |
| ↑ | ↑ | ↑ |
| → | → | → |
| ↓ | ↓ | ↓ |
| ↔ | ↔ | ↔ |
| ↵ | ↵ | ↵ |
| [| ⌈ | ⌈ |
|] | ⌉ | ⌉ |
| ⌊ | ⌊ | ⌊ |
| ⌋ | ⌋ | ⌋ |
| ◇ | ◊ | ◊ |
| ♠ | ♠ | ♠ |
| ♣ | ♣ | ♣ |
| ♥ | ♥ | ♥ |
| ♦ | ♦ | ♦ |