

## INFORMACIÓ I CODIFICACIÓ

### **Dades binàries**

Qualsevol dada que no siga text, es considera dada binària.

Per exemple: música, vídeo, imatge, un arxiu Excel, un programa, ...

La forma de codificar aquest tipus de dades a la seva forma binària és molt variable. per exemple en el cas de les imatges, cada punt (píxel) de la imatge es codifica utilitzant el seu nivell de vermell, verd i blau. De manera que una sola imatge produeix milions de dígit binaris.

En qualsevol cas siga quina siga la informació que estem codificant en binari, per poder accedir a aquesta informació, l'ordinador necessita el programari que sàpiga com descodificar la mateixa, és a dir saber què significa cada dígit binari per traduir a una forma més humana. Això només és possible utilitzant el mateix programari amb el qual es va codificar o bé un altre programari però que sigui capaç d'entendre la informació codificada.

### **Text**

El text és potser la forma més humana de representar informació. Abans de l'arribada l'ordinador, la informació es transmetia mitjançant documents o llibres en paper. Aquesta forma de transmetre és mil·lenària i segueix sent la forma més habitual de transmetre informació entre humans; fins i tot amb la tecnologia actual aplicacions com twitter, whatsapp, ... segueixen usant el text com a format fonamental per a transmetre informació.

Quant va aparèixer la informàtica com una ciència digital, va aparèixer també el problema de com codificar text en forma de dígit binaris per fer-ho representable a l'ordinador. La forma habitual ha estat codificar cada caràcter en una sèrie de nombres binaris. De manera que ara el caràcter A fora ara 01000001 i la B el 01.000.010....

El problema va sorgir per la manca d'estandardització, la lletra A es podia codificar diferent en diferents ordinadors i així ens trobàvem amb un problema en voler passar dades d'un ordinador a un altre. A poc a poc van aparèixer estàndards per intentar que tot el maquinari i programari codificarà els caràcters igual.

El codi ASCII

El problema de la codificació de text que feia incompatibles els documents de text entre diferents sistemes, es va pal·liar quan es va idear el 1967 un codi estàndard per part de l'ANSI, l'agència d'estàndards nord-americana, aquest codi és l'anomenat ASCII (American Standard Code for Information Interchange, codi estàndard americà per a l'intercanvi d'informació). El codi utilitza l'alfabet anglès (que utilitza caràcters llatins) i per codificar tots els possibles caràcters necessaris per a escriure en anglès es va idear un sistema de 7 bits (amb 7 bits es poden representar 128 símbols, suficients per a totes les lletres de l'alfabet anglès, en minúscules i majúscules, caràcters de puntuació, símbols especials i fins i tot símbols de control).

Tabla ASCII i binari, hexadecimal i decimal:

[http://es.wikipedia.org/wiki/American\\_Standard\\_Code\\_for\\_Information\\_Interchange](http://es.wikipedia.org/wiki/American_Standard_Code_for_Information_Interchange)

Taules de codis:

<http://www.microsoft.com/globaldev/reference/cphome.msp>

Però, en països amb llengües diferents de l'anglès, va sorgir el problema que part dels símbols de les seves alfabet quedaven fora de l'ASCII (com la lletra ñ) - Per això es van dissenyar codis de 8 bits

que afegien 128 símbols més i així van aparèixer els anomenats codis ASCII estesos. En ells, els 128 símbols primers són els mateixos de la taula ASCII original i els 128 següents es corresponen al símbol extra. Així per exemple el sistema MS-DOS utilitzava l'anomenat codi 437 que incloïa símbols i caràcters d'altres llengües d'Europa Occidental i caràcters que permetien fer marcs i vores en pantalles de text, entre d'altres símbols.

No obstant això 8 bits segueixen sent insuficients per codificar tots els alfabetes del planeta. Pel que cada zona usava la seva pròpia taula ASCII estesa. Davant el caos consegüent, la ISO va decidir normalitzar aquestes taules de codis per aconseguir versions estàndards dels mateixos.

Qué és Unicode?

<http://es.wikipedia.org/wiki/Unicode>

<http://msdn.microsoft.com/es-es/goglobal/bb688113.aspx>

Qué és UTF-8?

<http://es.wikipedia.org/wiki/UTF-8>

Consorti Unicode:

<http://www.unicode.org/>

Llista de caràcters Unicode (dels 65536 primers caràcters):

[http://en.wikibooks.org/wiki/Unicode/Character\\_reference/0000-0FFF](http://en.wikibooks.org/wiki/Unicode/Character_reference/0000-0FFF)

Representació d'alguns dels caràcters Unicode en codificació UTF-8 en hexadecimal:

<http://www.sslug.dk/~chlor/utf-8/>

## PRÀCTICA 1

Instal·lació del Visual Studio Code i canvi d'idioma de la interfície

## PRÀCTICA 2

Ejecuta en l'editor de text i escriu el següent text.

*Esta fila es sólo de caracteres ASCII?  
Esta línea té algún carácter no ASCII per començar?  
La eñe  
Cançó del Pingüi*

Guàrda'l en diferents formats i compara les grandàries dels arxius