

JAVA. LLEGIR DEL TECLAT EN WINDOWS

Els comentaris que segueixen a continuació, son per aquells usuaris que fan servir **Windows** i volen fer programes de consola que escriguin i llegeixin caràcters no anglesos, per executar des de dins o des de fora del **NetBeans**.

Per entendre el problema, anem a provar un petit programa que llegirà un text i l'escriurà a la consola. El programa complet s'ha indicat a continuació:

```

1 import java.util.Scanner;
2 import java.nio.charset.Charset;
3 public class Practiques {
4     public static void main(String[] args) {
5         Scanner sc = new Scanner(System.in);
6         System.out.println("Codificació que fa servir el programa: " + Charset.defaultCharset());
7         System.out.println("Copia i enganxa el text: ΑΕΙΟΥΑΒΕΙΟΎΑΒΕΙΌΎαεϊούάειϐύαιϐύαιϐύνŊçç€");
8         String llegit = sc.nextLine();
9         System.out.println("El tex és '" + llegit + "'");
10    }
11 }

```

Si l'executem des del **NetBeans** o bé des del sistema operatiu (en una consola **DOS**), obtindrem les sortides que s'indiquen a continuació:

- **NetBeans**

run:

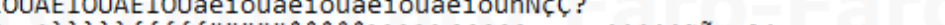
[illegible]

- **Consola DOS**

```
C:\Java\Practiques\build\classes>java Practiques
Codificació que fa servir el programa: windows-1252
Copia i enganxa el text: ÀÈÌÒÙÁÊÍÓÚÃÊÏÖÜÃÊÏÖÜàèìòùáêíóúæïöüâêïöüñÑçÇ?
ÀÈÌÒÙÁÊÍÓÚÃÊÏÖÜÃÊÏÖÜàèìòùáêíóúæïöüâêïöüñÑçÇ?
El tex és '·ÔPäëµ?Öäé?ÓØ??¶Òxâê???? ?;£?????f????¤¥???'
C:\Java\Practiques\build\classes>
```

El que es pot apreciar, és el que ens hem trobat moltes vegades, tot i que el missatge sembla que es llegeix be, quan entra dins del nostre programa i el tornem a escriure, la representació dels caràcters accentuats i símbols no anglesos no és la correcta. El que hem de saber és que no es pot resoldre pels dos entorns simultàniament, ja que la consola del **Windows**, que no utilitza el joc de caràcters del mode gràfic, fa servir la codificació **cp850**, mentre que el **NetBeans** fa servir un emulador de consola que utilitza la codificació configurada al **NetBeans** (per defecte **UTF-8**) per escriure, mentre que quan llegim text des de la consola de **NetBeans**, aquest s'introdueix seguint la codificació de **Windows cp1252**.

És interessant saber que **java** permet especificar la codificació desitjada en executar el codi (és un paràmetre de la crida, **-Dfile.encoding=codi_desitjat**), per exemple:



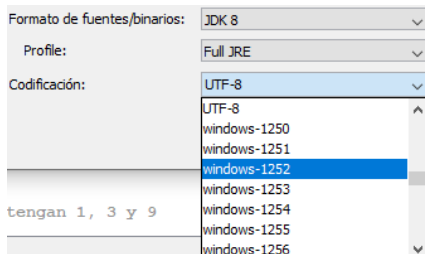
```

C:\Java\Practiques\build\classes>java -Dfile.encoding=cp850 Practiques
Codificació que fa servir el programa: IBM850
Copia i enganxa el text: ÀÈÌÒÙÁÉÍÓÚÂÊÏÔÛàèìòùáéíóúâêïôûñçç?
ÀÈÌÒÙÁÉÍÓÚÂÊÏÔÛàèìòùáéíóúâêïôûñçç?
El tex és 'ÀÈÌÒÙÁÉÍÓÚÂÊÏÔÛàèìòùáéíóúâêïôûñçç?'
```

D'aquesta forma, hem resolt el problema per una consola **DOS**, tot i que no completament, ja que el símbol del €, no s'ha reconegut en aquest sistema de codificació.

SOLUCIÓ PER A LA CONSOLA DE NETBEANS

Una primera forma de solucionar el problema per a **NetBeans**, és fer que el projecte faci servir la codificació del **Windows (cp1252)**, enlloc de la codificació per defecte que fa servir **NetBeans (UTF-8)**. Anem al menú **Archivo – Proyecto Properties**, seleccionem **Fuentes** i escollim la codificació desitjada pels fitxers entre les moltes disponibles i resolt.



run:

Codificació que fa servir el programa: windows-1252

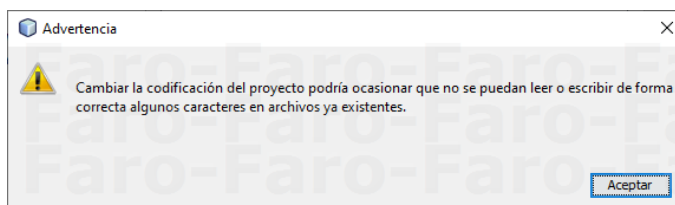
Copia i enganxa el text: ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ

ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ

El tex és 'ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ'

BUILD SUCCESSFUL (total time: 9 seconds)

Això té un petit inconvenient, del qual ens adverteix el mateix entorn, però si sempre programem en **Windows**, i no fem servir codi de tercers no tindrem cap problema.



Si no desitgem modificar el joc de caràcters que utilitza per defecte **NetBeans**, per exemple perquè treballem en **Windows** i **Linux** amb els mateixos projectes, i tampoc no ens volem recordar de cridar als nostres programes en **java** indicant la codificació que volem fer servir, tenim una altra alternativa que es comentarà a continuació. Per comprovar que funciona, torneu a deixar la configuració de caràcters de **NetBeans** a **UTF-8**, la que ve configurada per defecte.

Resolució del problema sense canviar la configuració de NetBeans

Ara que sabem una mica més de fitxers, sabem que l'ordre **Scanner** admet un segon paràmetre que és una cadena indicant la codificació del fitxer. Hem de saber que la codificació també es pot indicar quan llegim del teclat. L'únic que hem de fer és afegir la codificació desitjada a la línia:

```
public static void main(String[] args) {
    Scanner sc = new Scanner(System.in, "cp1252");
```

Ara, en executar el codi anterior des del **NetBeans**, obtindrem:

run:

Codificació que fa servir el programa: windows-1252

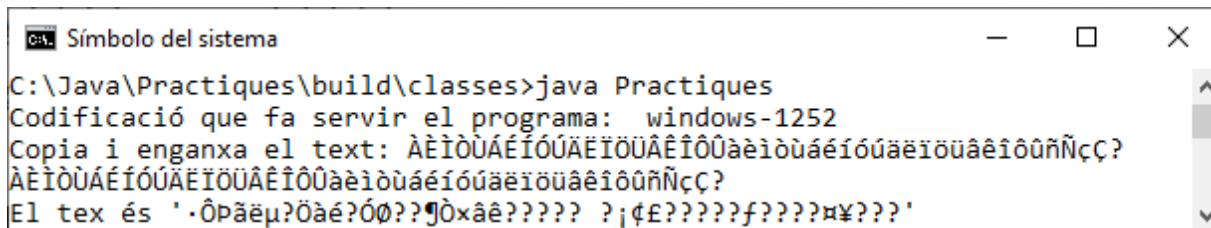
Copia i enganxa el text: ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ

ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ

El tex és 'ÀÈÌÒÙÁÊÍÓÚÃËÎÔÛàèìòùáêíóúãëïôûñÑçÇ'

BUILD SUCCESSFUL (total time: 8 seconds)

I des d'una consola DOS?



Evidentment, si executem el programa des d'una consola **DOS**, segueix sense funcionar, ja que la codificació **cp1252** només la utilitza **Windows** en el mode gràfic. És més, si ara executem el programa indicant-li a **java** que faci servir el sistema de codificació **cp850** (que ens havia funcionat abans), veurem que res ha canviat:

```
C:\Java\Practiques\build\classes>java -Dfile.encoding=cp850 Practiques
Codificació que fa servir el programa: IBM850
Copia i enganxa el text: ÀÈÌÒÙÁÉÍÓÚÂÊÏÔÜàèìòùáéíóúâêïôüññçç?
ÀÈÌÒÙÁÉÍÓÚÂÊÏÔÜàèìòùáéíóúâêïôüññçç?
El tex és '.ôpăëµ?öàé?óó??jòxâé???? ?;ıçf????f????ıç???'
C:\Java\Practiques\build\classes>
```

D'això deduïm que la codificació indicada a l'ordre **Scanner**, preval sobre la especificada en la crida a l'interpret de **java**.


Quina codificació fa servir la consola en Windows?

A la pàgina que s'indica a sota, ens parla de la pàgina de codis de la consola del **Windows**, que a Espanya acostuma a ser la **850**.

<https://docs.microsoft.com/es-es/windows-server/administration/windows-commands/chcp>

Ho podem esbrinar amb: `c:\Java\Practiques\build\classes>chcp`
(o amb **mode con**) `Página de códigos activa: 850`

La podem canviar (amb la mateixa ordre). Fins i tot podem indicar que faci servir el codi de caràcters del mode gràfic de **Windows**, el **cp1252**.



```
Símbolo del sistema
C:\Java\Practiques\build\classes>chcp
Página de códigos activa: 850

C:\Java\Practiques\build\classes>chcp 1252
Página de códigos activa: 1252
```

Si ara executem el programa anterior, on hem indicat a la funció **Scanner** que faci servir el codi **cp1252** de **Windows**, obtindrem la sortida que es mostra a continuació:

```
C:\Java\Practiques\build\classes>java Practiques
Codificació que fa servir el programa: windows-1252
Copia i enganxa el text: ÀÈÌÒÙÀÉÍÓÚÀÈÌÒÙàèìòùáéíóúáèìòùñÑçç€
ÀÈÌÒÙÀÉÍÓÚÀÈÌÒÙàèìòùáéíóúáèìòùñÑçç€
El tex és 'ÀÈÌÒÙÀÉÍÓÚÀÈÌÒÙàèìòùáéíóúáèìòùñÑçç€'
C:\Java\Practiques\build\classes>
```

On fins i tot el caràcter de € ha quedat adequadament representat. Malauradament, la modificació del joc de caràcters de la consola és transitori, només funciona mentre mantinguem la consola oberta (si executem una altra consola dintre d'aquesta també funcionarà en la nova), si la tanquem, tornarem a la pàgina de codis anterior (la 850).

Una altra possibilitat, és indicar-li a la classe **Scanner** que faci servir la codificació **850**:

```
public static void main(String[] args) {
    Scanner sc = new Scanner(System.in, "cp850");
```

```
C:\Java\Practiques\build\classes>java Practiques
Codificació que fa servir el programa: windows-1252
Copia i enganxa el text: ÀÈÌÒÙÀÉÍÖÙÀÆÏÖÙÀÆÏÖÙÀæïöùàæïöùàæïöùàæïöùàñÑçç?
ÀÈÌÒÙÀÉÍÖÙÀÆÏÖÙÀæïöùàæïöùàæïöùàæïöùàæïöùàñÑçç
El tex és 'ÀÈÌÒÙÀÉÍÖÙÀÆÏÖÙÀÆÏÖÙÀæïöùàæïöùàæïöùàæïöùàæïöùàñÑçç.'
```

Evidentment, tot i que la consola de **NetBeans** ens pot anar bé a l'hora de depurar els nostres programes per executar-los sense sortir de l'entorn, normalment ens interessarà més que el nostre programa funcioni correctament des de fora de l'entorn.

ESCRUIRE A LA CONSOLA INDICANT LA CODIFICACIÓ

Podem fer que el nostre programa **java**, escrigui amb un altre sistema de codificació que el que utilitza per defecte (per exemple per utilitzar emuladors de consola que fan servir un joc de caràcters diferent). Proveu aquest programa:

```

1  import java.nio.charset.Charset;
2  import java.io.*;
3  public class Practiques {
4      public static void main(String[] args) {
5          try{
6              String s = "ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ\n"; // Cadena que escriurem
7              System.out.println("Codificació que fa servir el programa: " + Charset.defaultCharset());
8              System.out.print("Escriptura normal: " + s);
9              OutputStreamWriter osw1 = new OutputStreamWriter(System.out,"cp1252");
10             osw1.write(s); // No escriu fins que s'omple el buffer o fem flush() o close()
11             osw1.flush(); // Si fem el close() a continuació, no seria necessari
12             OutputStreamWriter osw2 = new OutputStreamWriter(System.out,"utf-8");
13             osw2.write(s); // No escriu fins que s'omple el buffer o fem flush() o close()
14             osw2.flush(); // Si fem el close() a continuació, no seria necessari
15             OutputStreamWriter osw3 = new OutputStreamWriter(System.out,"cp850");
16             osw3.write(s); // No escriu fins que s'omple el buffer o fem flush() o close()
17             osw3.flush(); // Si fem el close() a continuació, no seria necessari
18             System.out.println("Escrivim normalment abans de tancar els fluxes");
19             osw1.close(); osw2.close(); osw3.close(); // Tanquem els fluxes
20             System.out.println("Després, l'escriptura podria no produir-se");
21         } catch (Exception e) {
22             System.out.println("ERROR: " + e.getMessage());
23         }
24     }
25 }

```

Provarem ara d'executar el programa des del **NetBeans** i des d'una consola de **DOS**.

- **NetBeans** (configurat per treballar en UTF-8)

run:

Codificació que fa servir el programa: UTF-8

Escriptura normal: ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ

ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ

ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ

ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ

Escrivim normalment abans de tancar els fluxes

BUILD SUCCESSFUL (total time: 0 seconds)

- **Consola DOS**

```

C:\Java\Practiques\build\classes>java Practiques
Codificació que fa servir el programa: windows-1252
Escriptura normal: ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ?
ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ?
ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ?
ÀÈÌÒÙÀÉÍÓÚÀËÏÖÜàëìòùáéíóúäëïöüñŇčĚ?
Escrivim normalment abans de tancar els fluxes
C:\Java\Practiques\build\classes>

```

El funcionament en els dos cassos és molt semblant. Curiosament, en escriure normalment amb **System.out.println**, ens adaptem a la configuració de la consola utilitzada (**DOS** o **NetBeans**). Fixeu-vos que quan executem el codi des de **NetBeans** el programa indica que utilitza **UTF-8**, però des de la consola diu que es fa servir **cp1252**. Tanmateix, si escrivim a la consola com si es tractés d'un fitxer, només quan fem servir la codificació adequada (**cp850** al **DOS** i **UTF-8** al **NetBeans**), el missatge es representarà correctament.

