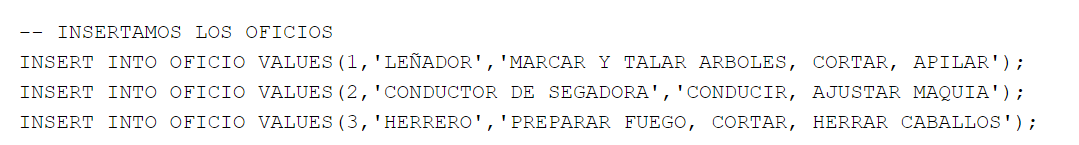
Descripción del Caso

En Drive tenemos un fichero zip con el esquema de base de datos referido en la página 21 del tema3 del módulo de base de datos (ver [enlace](https://drive.google.com/drive/folders/1ocEjY7SUf2OtAnWAYygXHvwQddksXHvy))

Dentro del fichero comprimido hay una serie de ficheros sql, el fichero esquema.sql no sólo crea el esquema de BD sino que también inserta registros en las tablas, vamos a centrarnos en el registro que pone "LEÑADOR" para la tabla oficio



En muchas ocasiones verás que esa Ñ no se visualiza bien desde sqlplus y/o developer, antes de proponerte una solución es necesario saber lo que ocurre.

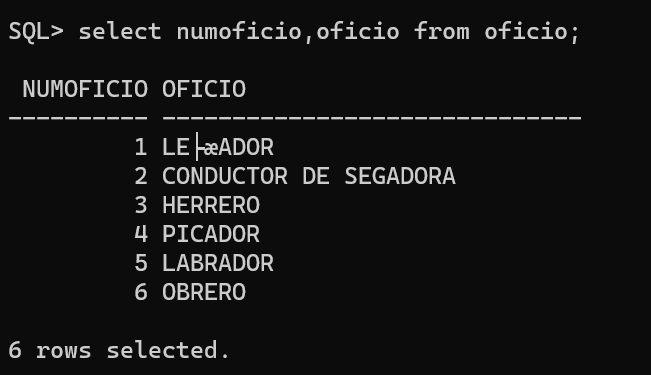
Esa Ñ se guarda con 0 y 1 siguiendo un tipo de codificación, hay varias codificaciones y no todas usan la misma secuencia de 0 y 1, simplificando mucho te diré que el origen del problema radica en que la codificación del cliente no coincide con la de del servidor.

Explico un poco a qué me refiero con cliente/servidor:

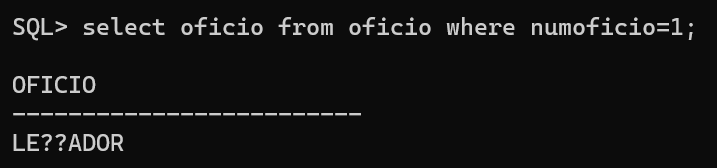
* Por una parte el cliente (sqlplus, developer etc) que es con el que introducimos**/visualizamos esos datos.**
* Por otra parte del servidor (base de datos) donde **grabaremos esos datos** según una codificación concreta

Vamos a ver ejemplos en los que se visualiza mal la Ñ, esa letra podría estar bien o mal grabada en la base de datos (ya explicaré más adelante qué quiero decir)

CASO 1) Se visualizan otros caracteres en lugar de la Ñ, el cliente realiza una interpretación que no corresponde con la Ñ que pusimos originalmente.



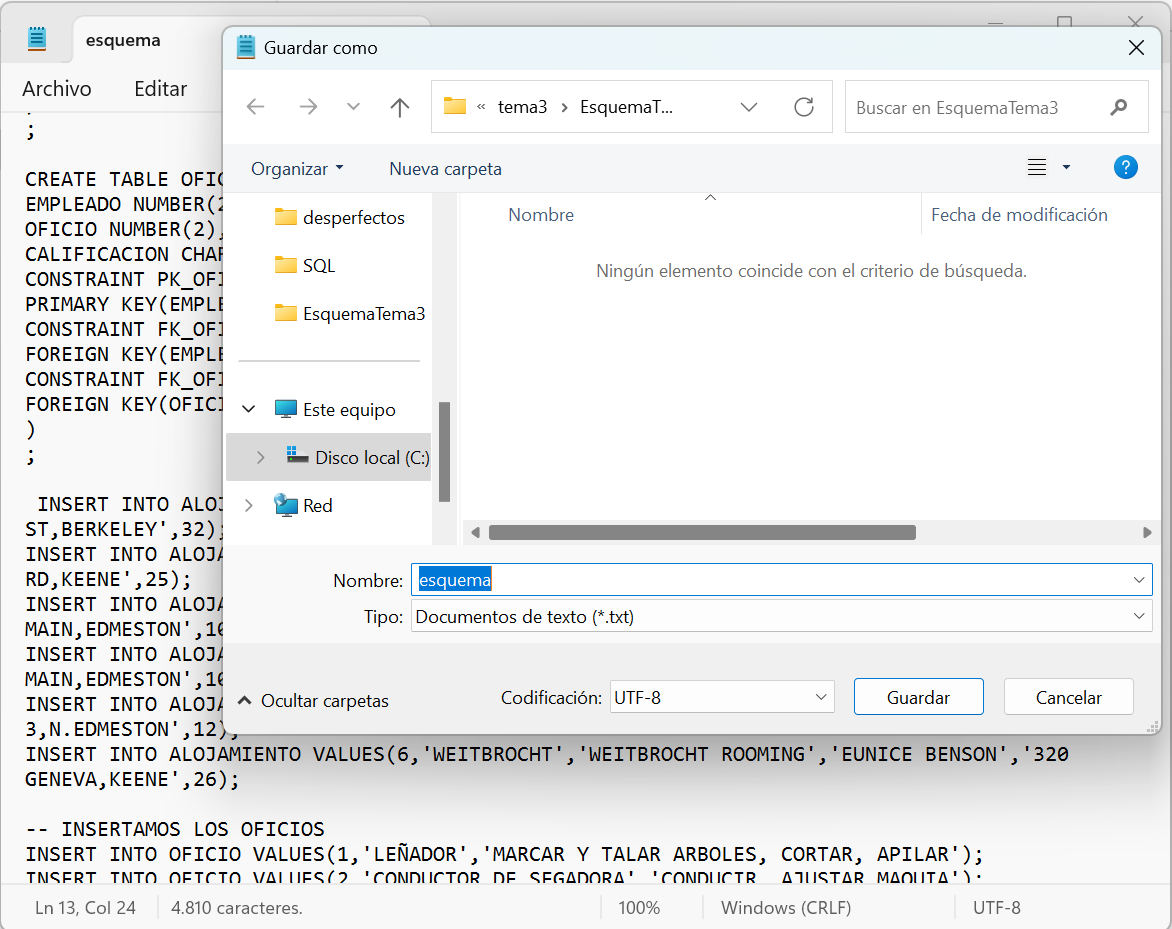
CASO 2) Obtenemos unos ? ya que los 0 y 1 recuperados no se pueden visualizar con la codificación activa. Cada ? representa un carácter que no se ha podido representar.



Codificación de caracteres

Tenemos muchas formas de codificar esa Ñ, nuestro sistema operativo conoce la codificación usada en ese fichero y lo normal es que lo muestre correctamente cuando usas una aplicación desde el entorno gráfico.

**¿Cómo vemos esa codificación?** Una forma sencilla es editar el fichero con el notepad y fijarnos en la codificación cuando pulsamos en guardar como



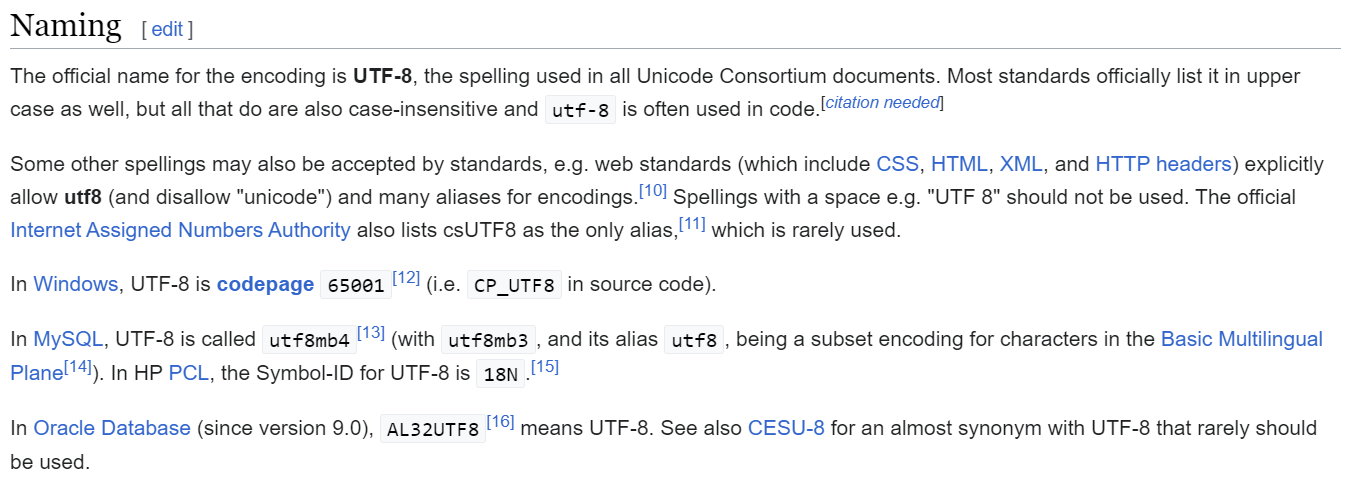
Si te fijas la codificación es UTF-8, si echas un vistazo en la Wiki para esa codificación verás que es el Santo Grial de las codificaciones, y será nuestra codificación de referencia para representar nuestra información de texto tanto en base de datos como en programación (ver [enlace](https://en.wikipedia.org/wiki/UTF-8))

Imaginarás que una de las formas que todo vaya bien, todos los elementos involucrados (los llamo **sitios** para simplificar), intérprete de comandos de windows, sqldeveloper etc y sin olvidarnos del más importante, la propia base de datos, almacenen esa Ñ en la misma codificación, en el mejor de los casos se grabará bien pero en algunos sitios se verá mal y en el peor se grabará mal y no se verá bien en casi ninguno.

**NOTA:** No es cierto que cliente (sqlplus, developer etc) deban usar la misma codificación que el servidor (base de datos). Oracle establece mecanismos para realizar las traducciones que se necesiten pero siempre que se configure adecuadamente.

Tampoco es cierto que si cliente y servidor usan la misma codificación no se produzcan problemas ya que existen una serie de parámetros que le informan a Oracle de la codificación usada en el cliente y podemos configurarlo de forma incorrecta

Si sigues leyendo la wiki verás una referencia de la denominación que hace cada sitio de la codificación utf8, pego una captura de pantalla.



Cuando se refiere a windows lo hace a la codificación usada en el cmd, cada aplicación gráfica suele manejar correctamente utf8 y permite configurar fácilmente la codificación usada (te recuerdo que ya lo haces cuando programas usando Eclipse)

NLS\_LANG (¿Qué es?)

Esta variable permite la configuración del Idioma/región/juego de caracteres usado en el cliente.

NLS viene de National Language Support.

Precedido de NLS vendrá una serie de parámetros que no sólo pondremos en el cliente sino también en la base de datos y que determina el comportamiento de la base de datos frente a aspectos del idioma como manejo de fechas etc y el más importante, el juego de caracteres.

El parámetro CHARACTERSET (**NLS\_CHARACTERSET**), uno de los más importante, aparecerá en la configuración de:

* **La sesión** , indicando el juego de caracteres que se sabe usa el cliente y que usaremos durante esa sesión en la que nos conectamos a la base de datos. Normalmente se indicará configurando la variable NLS\_LANG
* **Servidor**, indica qué juego usa la base de datos para grabar esos caracteres.

Esta configuración NLS se puede realizar en varios niveles:

* Instancia (conjunto de procesos que hacen que funcione ese SGBD)
* Base de datos (los datos de la base de datos)
* Sesión (La sesión del cliente)

En cada uno de estos niveles tendremos configurados una serie de parámetros, los de sesión se configuran normalmente en función de la configuración que pongamos en **NLS\_LANG**, ante su ausencia se hace por valores predeterminados.

La configuración de Sesión prevalece sobre la de instancia y ésta sobre la de Base de datos.

Para ver éstos parámetros prueba a ejecutar lo siguiente:

**column parameter format A30**

**column value format A30**

**select \* from nls\_session\_parameters;**

**select \* from nls\_database\_parameters;**

**select \* from nls\_instance\_parameters;**

Algunos de los parámetros de instancia y base de datos se pueden cambiar de forma sencilla pero otros son más traumáticos como el cambio de la codificación de la base de datos **NLS\_CHARACTERSET** que se configura al crear la base de datos.

Desde una misma sesión se pueden modificar algunos de esos parámetros de sesión con el comando **ALTER SESSION**

**Con NLS\_LANG estamos diciendo a Oracle no sólo nuestro idioma sino también el juego de caracteres que usamos para la manipulación/visualización de los datos de la base de datos desde el cliente por lo que una mala configuración puede hacer que no resolvamos este problema de la Ñ.**

**No estamos obligados a poner el mismo juego de caracteres en cliente y servidor ya que Oracle realizará las traducciones necesarias para que se vea bien (juego de caracteres cliente) pero se grabe en el juego de caracteres que usa la base de datos.**

Si queremos hacer bien las cosas en el cliente tendremos una primera fase de configuración que ya tendremos resuelta y consistente en:

* Configurar el idioma (por ejemplo instalando windows en Español)
* Configuración regional (personalizando elementos como que la , es separador decimal)

La segunda fase ya dependerá del tipo de cliente que usemos y consistirá en:

* Configurar juego de caracteres en el cliente que usemos (escogiendo un juego de caracteres que permita representar los caracteres del idioma, en el caso de cmd se hace con el comando chcp)
* Guardando las instrucciones (select , insert) en un juego de caracteres que permita representar esos caracteres del idioma

**NOTA:** Aquí tendrás que tener cuidado porque en especial en windows, el juego de caracteres usado en el fichero seguramente no coincida con el juego de caracteres usado en el cliente (por ejemplo cmd) y ya tenemos el lío

**La tercera y más importante consistirá en informar a Oracle de todas éstas configuraciones previas y eso lo haremos configurando adecuadamente NLS\_LANG en nuestro cliente pero siento decirte que no todos los clientes usan NLS\_LANG.**

**NOTA:** No tiene ningún sentido que configures windows en Alemán y luego te conectes con una configuración NLS\_LANG en español, por ello la configuración del sitio desde el que te conectas debería ser compatible con la configuración que pongas en NLS\_LANG

Ahora toca saber

* ¿Qué valores se pueden poner en NLS\_LANG? y ¿Dónde se configura?
* ¿Cómo puedo saber/cambiar la codificación que usa cada sitio?

¿Qué valores se pueden poner en NLS\_LANG?

En NLS\_LANG se puede configurar:

* **Language (Idioma)** Para cuestiones como que salgan los mensajes en español, nombres de los días y meses etc. Si no se pone se asume AMERICAN
* **Territory (Región)** Para formato de fecha, de número, de monedas, en caso de no definirlo se usa la configuración del Idioma.
* **Charset (juego de caracteres)** Juego de caracteres, ten cuidado porque cada idioma tiene un juego de caracteres por defecto por lo que no tiene por qué configurarse en el juego de la base de datos

Si no configuras NLS\_LANG, la configuración por defecto será **AMERICAN\_AMERICA.US7ASCII**

Puedes ver esa tabla de codificación (us7ascii) en el siguiente enlace (ver [enlace](https://www.ascii-code.com/es))

Todos son opcionales pero lo mejor es poner cada uno de ellos, la forma de separar cada uno es con los siguientes caracteres (te pongo un ejemplo para que lo veas)

**NLS\_LANG=SPANISH\_SPAIN.AL32UTF8**

No te puedes inventar los valores por lo que tendremos que saber qué se puede poner y qué no, usa la vista siguiente para ver esos valores

**column parameter format A32**

**column value format A32**

**select parameter,value from V$NLS\_VALID\_VALUES;**

Es importante entender que la configuración bien Idioma y territorio no solucionas nada si no pones un juego de caracteres adecuado:

* Que se ajuste al juego de caracteres que usa el cliente
* Que represente correctamente los caracteres del idioma

Vuelvo a insistir en el hecho que Oracle traducirá el juego de caracteres del cliente hacia/desde el servidor para que se visualice de forma correcta pero lo grabará en el juego de caracteres de la Base de datos.

**CUIDADO:** Si ambos coinciden (conf en NLS\_LANG y en Base de datos) no realizará conversión y lo grabará tal cual se lo demos

¿Dónde se configura NLS\_LANG? Caso cmd de windows

Ten cuidado porque aunque existan esos parámetros de sesión no siempre se configuran vía variable NLS\_LANG, ten presente que hay muchas formas de conectarte a la base de datos por ejemplo las conexiones Thin JDBC.

Ya que el problema ha surgido en el cmd veamos cómo hacer para el caso de windows

En windows se puede configurar de forma permanente esta variable en el registro del sistema o en una variable de entorno, por supuesto puedes probar a realizar cambios sobre la marcha desde el cmd para ver cómo afecta al comportamiento.

Una vez configurada la variable como var de entorno, podemos ver su valor de la siguiente forma

Desde cmd con **echo %NLS\_LANG%**

Desde sqlplus con **SQL> host echo %NLS\_LANG%**

Realmente son el mismo comando, lo que hace **host** al ponerlo en sqlplus es permitir la ejecución del comando siguiente en un cmd

Si quieres modificar el valor sobre la marcha en cmd no tienes más que poner

**set NLS\_LANG=<valor>**

Algunos de esos parámetros que podemos ver con nls\_session\_parameters, son accesibles desde la función userenv, puedes probar

**select userenv('language') from dual;**

¿Qué codificación usa cada sitio y cómo se puede cambiar?

Ya vimos que no tiene sentido configurar un juego de caracteres en NLS\_LANG diferente del que usemos en el sitio desde que nos conectemos a la base de datos por lo que lo primero es conocer la forma de ver la codificación usada y cómo cambiarla.

Aprovecho en este punto para ver también otros sitios donde pueden estar esas Ñ.

A modo de ejemplo veremos los tres sitios en los que manipularemos esos caracteres.

**Windows (cmd)**

Para saber la codificación se usa el comando **chcp** (ver [enlace](https://learn.microsoft.com/es-es/windows-server/administration/windows-commands/chcp))

Aquí tenemos dos casos según la versión de windows

En el caso de windows 10 lo normal es que aparezca la codificación 850 (no es utf8)

En el caso de windows 11 aparecerá 65001 que ya sabes es utf8

La codificación 850 se corresponde a WE8PC850 en Oracle

La codificación 65001 (utf8) se corresponde a AL32UTF8 en Oracle.

Daría como resultado las configuraciones de NLS\_LANG

* **SPANISH\_SPAIN.WE8PC850**
* **SPANISH\_SPAIN.AL32UTF8**

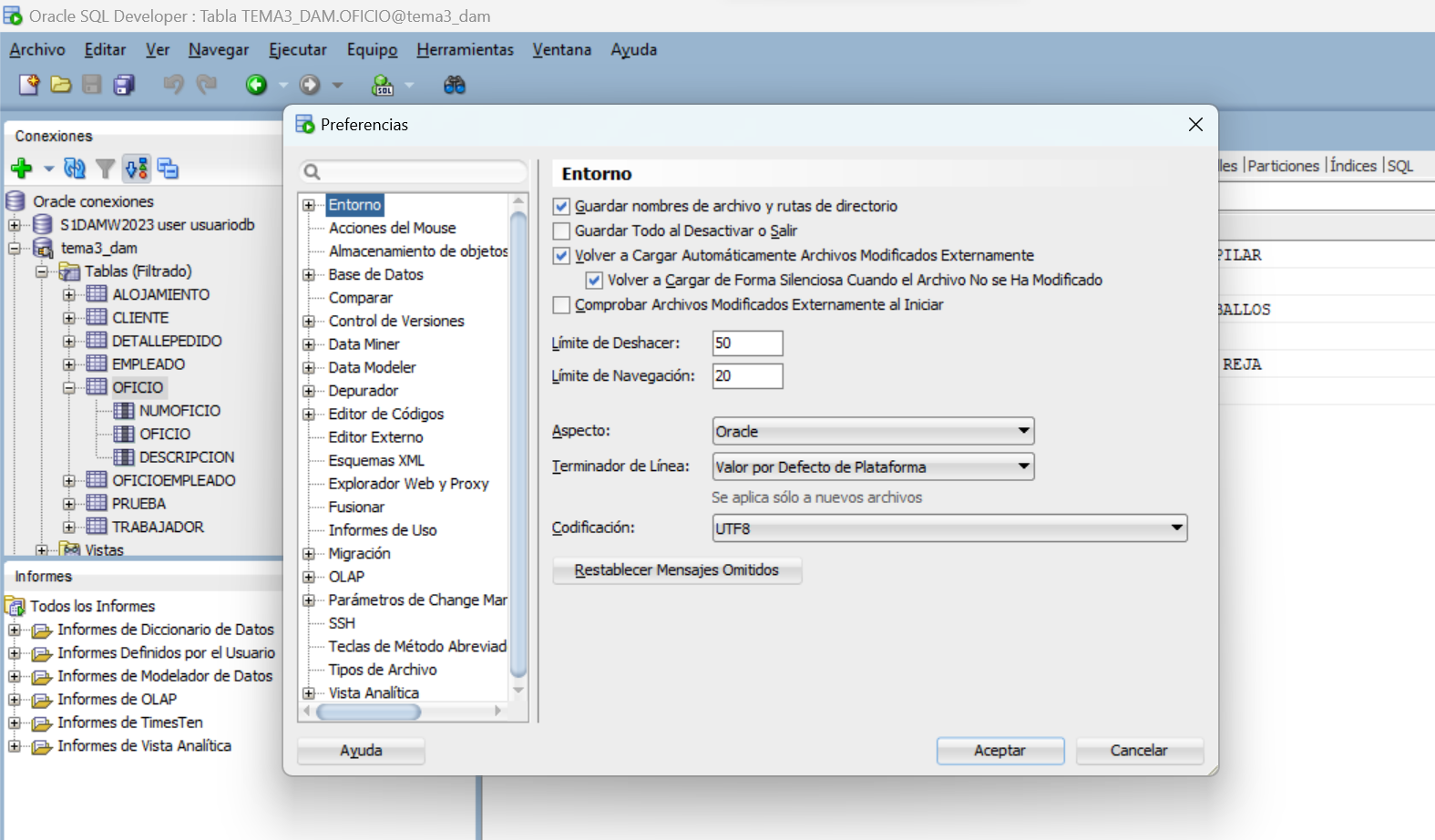
**Editor de fichero (ejemplo Notepad)**

Ya has visto al principio que suele ser la opción por defecto de la codificación de los entornos gráficos pero no está de más comprobar que el fichero sql en cuestión está en uft8, sino cambia la codificación desde guardar como.

NOTA: Hay quejas entre los usuarios de windows 11 ya que intentan cambiar la codificación de sus archivos desde notepad/notepad++ y no lo hace, se queda siempre en formato utf8

**SqlDeveloper**

Para ver la codificación (por defecto es utf8) puedes hacerlo desde Herramientas/preferencia/entorno

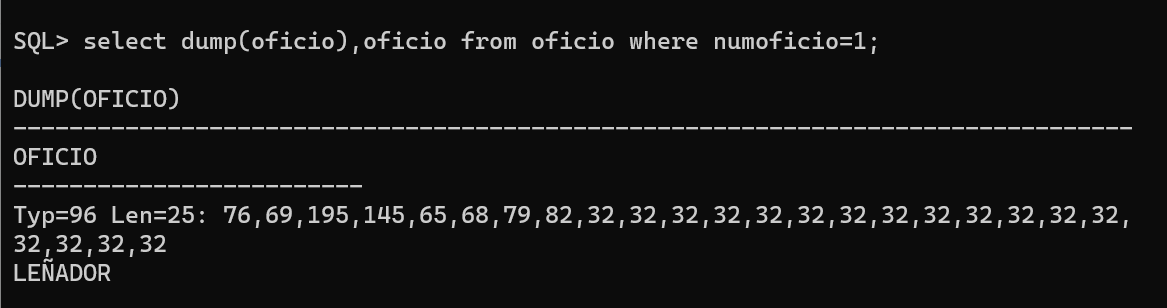


¿Cómo ver la representación interna de un dato en la base de datos?

Es muy interesante saber cómo se ha grabado un string en la base de datos porque resuelve la duda de si el problema de visualización deriva de una mala configuración o que se haya grabado de forma incorrecta.

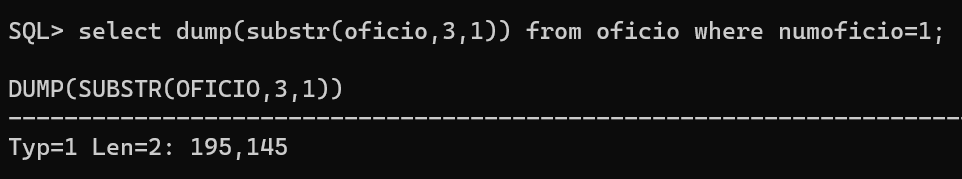
El comando dump permite ver esa información, prueba lo siguiente para ver cómo se ha grabado el string 'LEÑADOR' en la base de datos

**select dump(oficio),oficio from oficio where numoficio=1;**



Podemos afinar más y que nos muestre sólo la codificación de esa Ñ con

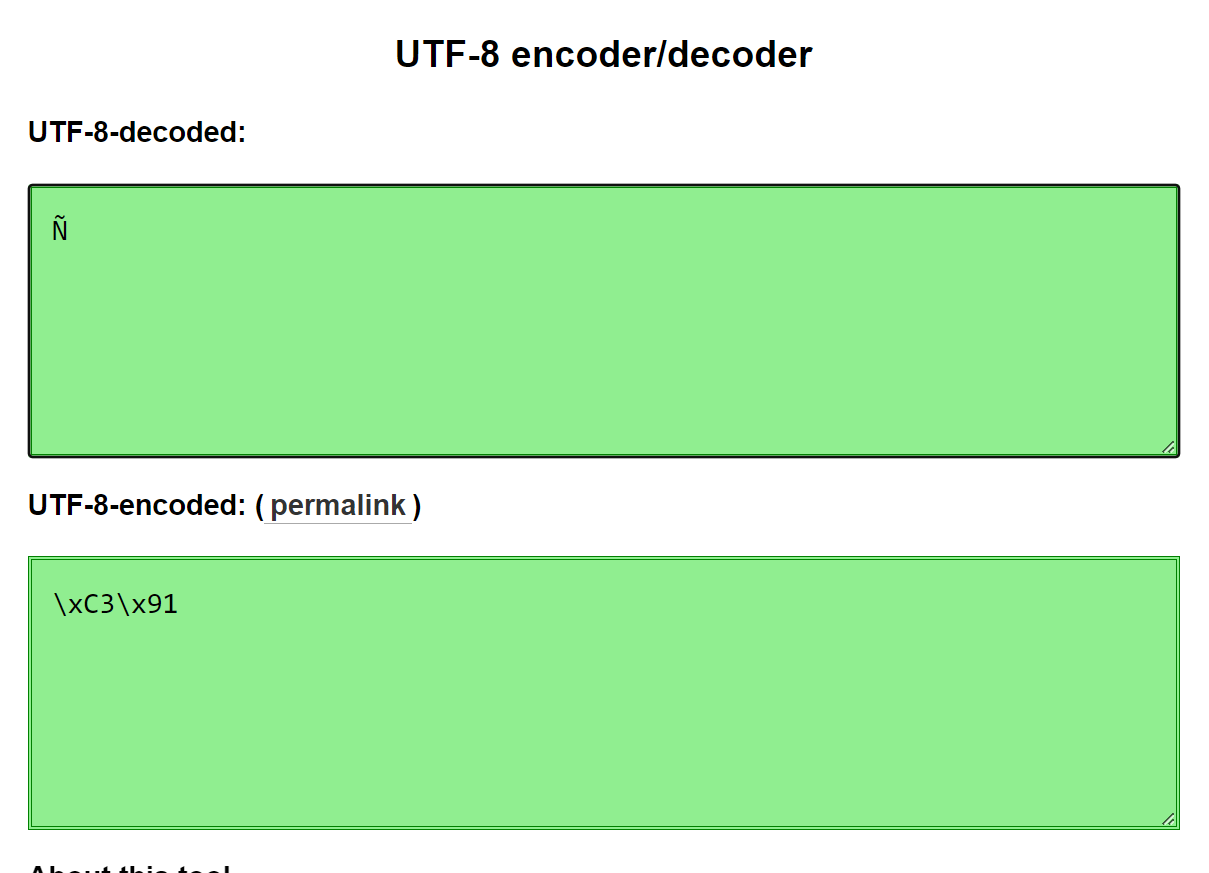
**select dump(substr(oficio,3,1)) from oficio where numoficio=1;**

****

Usaremos alguna página web que nos muestre la codificación de ese String en la base de datos siguiendo la codificación con la que se creó, en este caso UTF8

Puedes probar con la siguiente página <https://mothereff.in/utf-8>

Si pones el string y pulsas en codificar, te mostrará la codificación en formato Hexadecimal



La salida es \xC3\x91 que pasado a decimal es 195,145 y que coincide con lo grabado en la BD

Ahora vamos a hacer las cosas mal para ver un caso en el que se graban mal los datos en la base de datos, sigue los siguientes pasos:

Configuras NLS\_LANG=.WE8PC850 y pones chcp 850 en el cmd

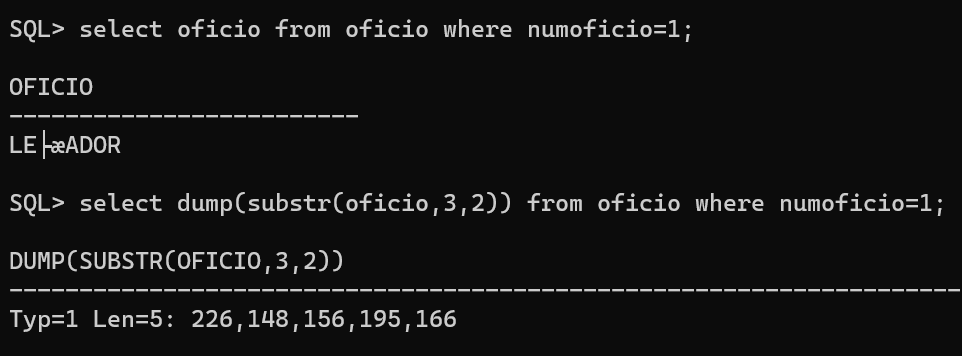
**NOTA:** la página de código 850 se asigna a NLS\_LANG con WE8PC850

**set NLS\_LANG=.WE8PC850**

**chcp 850**

Nos conectamos con sqlplus y ejecutamos lanzar.sql que está en UTF8

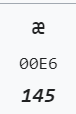
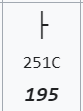
Seleccionamos el nombre del oficio que tiene la Ñ y pedimos el volcado de lo que se ha grabado en la base de datos, se visualiza lo siguiente



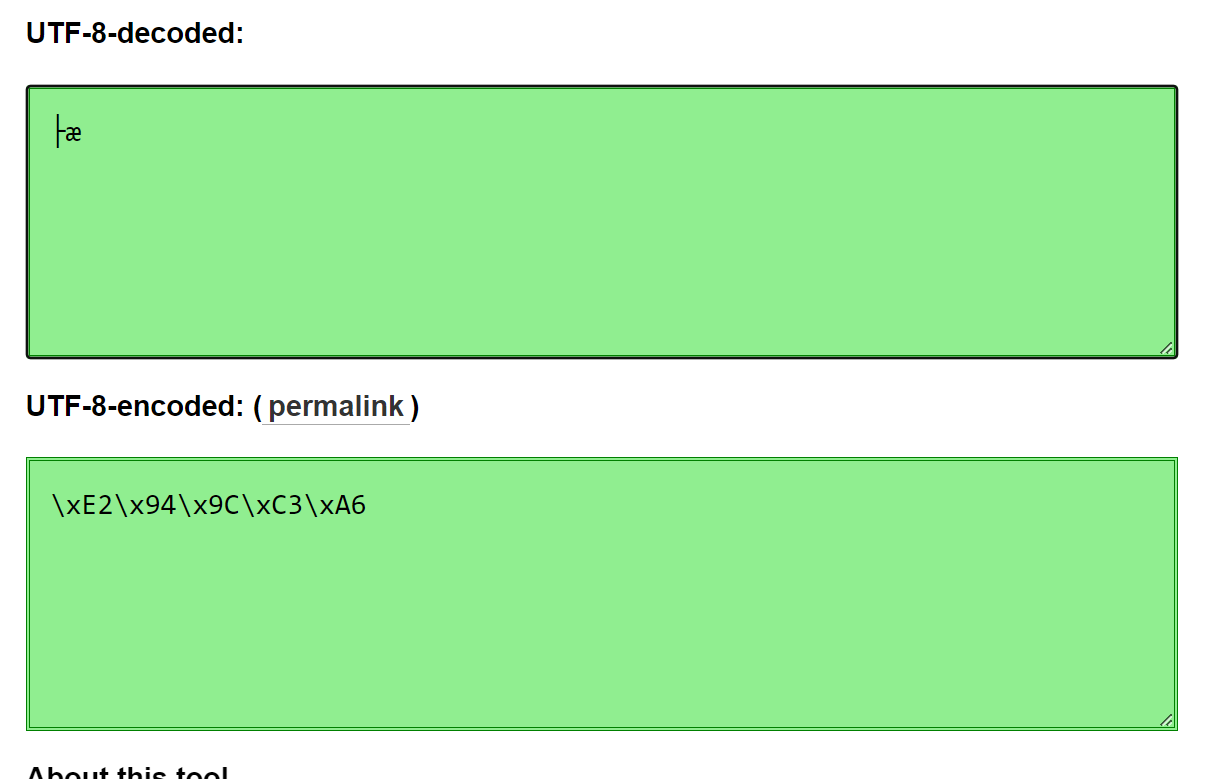
**¿Qué ha pasado?** Ahora no es que se visualice mal sino que se muestran otros caracteres diferentes a la Ñ

Lo que ha ocurrido es que Oracle interpreta que todo lo que recibe está en codificación 850 por lo que la codificación interna de la Ñ en utf8 (así está grabado en el fichero) , se interpreta siguiendo la tabla 850 (ver [enlace](https://es.wikipedia.org/wiki/P%C3%A1gina_de_c%C3%B3digos_850))

Según esa tabla la codificación interna de la Ñ (en Utf8) se interpreta como los caracteres siguientes



Esos caracteres se traducirán a UTF8 que es la codificación que usa la base de datos, si pones esos caracteres en la web <https://mothereff.in/utf-8> (copia/pega desde la salida de sqlplus) para que muestre la codificación interna, nos dirá que se grabará en la BD como:



Si pasamos esos números hexadecimales a decimal obtendremos:

226,148,156,195,166

ERRORES

**ERROR 1 (06-02-23):** He realizado la siguiente configuración en un windows 10 desde el cmd

**chcp 65001**

**set NLS\_LANG=SPANISH\_SPAIN.AL32UTF8**

Después me he conectado con sqlplus a la base de datos y la Ñ se visualizaba correctamente pero al escribir cualquier comando que incluyera la Ñ aparece de forma correcta pero se interrumpe la interpretación del comando en cuando llega a la Ñ.

select \* from oficio where oficio='LEÑADOR';

Si lo pruebo desde Windows 11 no tengo problema, en este caso chcp está configurado como 65001 por defecto (por lo menos en mi equipo)

NO TENGO CLARO SI AÑADIRLO

Puedes también usar un comando interesante que te permite convertir literales de un juego de caracteres a otro, puedes probar el siguiente código

>>>> No se si añadir los elementos siguientes

>>> Poner comando convert

>>>> No se si es interesante poner un dump que muestre cómo se grabaría la Ñ pero preguntando a dual.

Cuando no configuramos el characterset pero si idioma y regional no aparece NLS\_CHARACTERSET para sesion pero si se puede ver desde

select sid,serial#,client\_charset from v$session\_connect\_info;

Habrá que identificar la sesión propia con sid y serial

v$session\_connect\_info muestra las conexiones de red a la base de datos, client\_characterset sólo vale para Oracle >=11.1 y que no sean thin jdbc

Mirando esa tabla se ve que el charset por defecto es WE8ISO8859P1

Ese nombre tiene como descripción en Oracle = Western European 8-bit ISO 8859 Part 1

En windows corresponde a = ISO 8859-1 Latin 1; Western European (ISO)

Y el número de chcp es el 28591 (ver [enlace](https://learn.microsoft.com/en-us/windows/win32/intl/code-page-identifiers))

Los número chcp que nos podemos encontrar normalmente 850 y 65001, otros códigos que incluyan ñ podrían ser 1252 y 28591

Leer ésto bien y añadir lo que proceda

SQL> SELECT DUMP(col,1016)FROM tabla;

Tipo=1 Len=39 Conjuntodecaracteres=UTF8: 227,131,143,227,131,170

El 1016 se refiere al formato usado en la salida, 1000 para que ponga codificación y 16 para que la salida sea en decimal

¿Dónde se realiza la conversión de caracteres?

Normalmente, la conversión se realiza en el lado del cliente por razones de rendimiento. Así es desde la versión 8.0.4 en adelante. Si la base de datos está utilizando un juego de caracteres desconocido por el cliente, la conversión se realiza en el lado del servidor. Así es desde la versión 8.1.6 en adelante.

¿Por qué no muestra Windows SQL\*Plus todos mis caracteres extendidos?

Si ves cuadrados negros en lugar de caracteres, probablemente no has definido la fuente adecuada para tu página de códigos. Una fuente es una colección de glifos (de "jeroglíficos") que comparten una apariencia común (tipo de letra, tamaño de caracteres). El sistema operativo utiliza una fuente para convertir un valor numérico en una representación gráfica en la pantalla. Una fuente no contiene necesariamente una representación gráfica para todos los valores numéricos definidos en la página de códigos que está utilizando. Por eso a veces aparecen cuadrados negros en la pantalla si cambia las fuentes y la nueva fuente no tiene ninguna representación para un determinado símbolo.

La función de "mapa de juego de caracteres" en Windows se puede utilizar para ver qué glifos son parte de una fuente determinada.

Puede ser interesante ver las codificaciones en windows con la aplicación **charmap**

[http://www.unicode.org](https://www.unicode.org/)

[http://www.iso.org](https://www.iso.org/home.html)

[http://czyborra.com/charsets/iso8859.html](https://czyborra.com/charsets/iso8859.html)