Information about Codepages

German version of this page

General

A normal character set contains only 256 characters. 32 of them are used as control characters as Line Feed, Carriage Return and so on, so that only 224 are usable. Even restricted to latin characters, 224 characters are by far not enough for all language in the world. And don't forget the block graphic characters…
The solution is to use different character sets for different languages. A character set is called 'Codepage'.

Examples:

In english there are no special characters, so the createors of the first character set for IBM PCs added a lot of block graphic characters. This set is called codepage 437 today.
But it contains special characters too for german, french and others, like 'ß', 'ü', 'é'…

Codepage 865 for Norway and Denmark is nearly identical to 437. Only two block graphic characters was replaced by 'ø' and 'Ø'.

Codepage 850 has less block graphic characters and more special characters used in Europe, like 'ã', 'Ð', 'Þ'.

Codepage 852 is for slav languages and contains characters.

In codepage 862 there are 27 hebrew characters that replaces most of accent characters and all german umlauts of codepage 437.

Codepage 866 is for cyrillic.

The lower half of these codepages is identical (ASCII-Code).

Under english versions of Windows codepage 1252 is default. Because Windows is has a graphical interface, block graphic characters aren't needed. So, in codepage 1252 is a lot of space for special characters that are spreaded to serval codepages in DOS. That's why I could use all these different characters here…
Some characters for slav languages are in codepage 1250 only. For cyrillc there is codepage 1251, for greek 1253, for hebrew 1255, arabic 1256, and so on…

The list of codepages is long. Here is a overview:
http://www.kostis.net/charsets
Here are general information about 'Characters and encodings':
http://www.cs.tut.fi/~jkorpela/chars/index.html

Codepages under MS-DOS

Under MS-DOS the character set comes initially from the BIOS of the graphic card which contains codepage 437.

With introducton of EGA cards the character set became programmable. MS-DOS uses this feature for loading different codepages thru the driver DISPLAY.SYS. On the most of local versions of DOS and Windows 9x for Central Europe DISPLAY.SYS is loaded by default in the CONFIG.SYS:

```
C:\WINDOWS\COMMAND\display.sys con=(ega,,1)
```

The actual loading of a codepage is made by the MODE command. These two lines are default in german versions of DOS and Windows 9x. They prepare and activate codepage 850:

```
mode con codepage prepare=((850) C:\WINDOWS\COMMAND\ega.cpi)
mode con codepage select=850
```

The character set (the font) is loaded from the file EGA.CPI.
There are lots of programs that are either old or simply developed by users of codepage 437… This makes

trouble if another codepage than 437 is used. Have a look at this screenshot of the Norton Commander 2.0 with codepage 850.

You can create your one CPI-Files using the CPI tools from Kosta Kostis.

Codepage 850 has been extended by the new Euro currency symbol. Some sources call it codepage 858 then.

Some DOS applicaton uses the programmability of the character set to simulate a nice graphical interface in text mode. The good old PC-Tools 7.0 and higher from Central Point or the Norton Utilities 5.0 and higher from Symantec are good examples.
To program the character set works in pure DOS or DOS fullscreen mode only. In the DOS box always a Windows font is used. So this kind of programs looks strange in the DOS-Box. You can either disable the 'pseudo graphics', switch to fullscreen or you create a special font for the DOS-Box…

DOS Codepages under Windows

Under Windows MS-DOS and console applications runs either in fullscreen mode or in a window called the DOS-Box.
For both the codepage must be set separately because different mechanisms are used:

In the fullscreen mode the MS-DOS that runs under the hood of Windows 95/98 looks thru. Here applys the same as described above in section 'MS-DOS'.

Under NT, 2000 and XP the codepage is set by the command CHCP (Change Codepage) which changes the codepage for screen and keyboard. To work NLSFUNC must be loaded before.
Using the MODE CON command you can change the codepage for the screen separately but for a strange reason it does not work in the AUTOEXEC.NT.
Both work in fullscreen mode and in the DOS-Box if you are using the True Type font 'Lucida Console'.

In the DOS-Box there are two different types of fonts available: The above-mentioned True Type Font 'Lucida Console' and bitmap fonts named 'Terminal'.
Unter Windows 9x the latter are called 'Bitmap Fonts', under NT, 2000 and XP 'Raster Fonts'. Two names for exactly the same…

'Lucida Console' is a Unicode font and contains characters for all available codepages. Even hebrew characters are in it, but they arn't usable because of a missing table for codepage 862 in the file. So CHCP 862 doesn't work even on hebrew versions of Windows. To get codepage 862 in the DOS-Box of Windows 2000/XP you have to look for font files of Win9x called APP862.FON or VGA862.FON. I've created a single font size (10x19) with codepage 862 just for fun. But it's not reviewed so far: New862.ZIP

The 'Bitmap Fonts' or 'Raster Fonts' resp. must be installed fitting for the used codepage. For Windows 95 there is the program ChangeCP on the CD and it works fine. Windows 98 has a new one and I had no success with it. Under Windows ME there is the program MSCONFIG, section 'International' for changing the codepage.
Under NT, 2000 und XP you have to do it by hand. CHCP has no effect on the font when an 'Raster Font' is used.
You can install one of my fonts with the right codepage (and use a font size from the file).

Because the 'Bitmap Fonts' or 'Raster Fonts' have the correct codepage virtually accidental, you could use different codepages at the same time by installing fonts with different codepages and using the according font size.
Caution: It seems as Windows could handle multiple fonts for the DOS-Box with the same size but it cannot!
In the font preview it uses the diffenent fonts but in the DOS-Box it uses always the same…

Last change: 12th Feb. 2007

Uwe Sieber

eMail

Back