Understanding Coded Character Set ID (CCSID)

Craig Pelkie
Bits & Bytes Programming, Inc.
craig@web400.com

Understanding Coded Character Set ID (CCSID)

Edition CCSID_20020219

Published by

Bits & Bytes Programming, Inc.
Valley Center, CA 92082
craig@web400.com

Understanding Coded Character Set ID (CCSID)

When you transfer a file from the AS/400 to your PC, you expect the data to be converted from AS/400 EBCDIC to PC ASCII. The conversion is done by mapping characters from one character set to another. You can then view and use the data on the PC. However, some files on the AS/400 are not automatically translated. When you view those files, you see long sequences of hexadecimal data.

The AS/400 has an extensive set of attributes that control how conversions are done. Because the AS/400 is used around the world, OS/400 provides support for many national languages. The characters used in those languages vary widely. Accent marks on Spanish and French characters are some examples of different character sets that you may be familiar with.

To deal with these requirements, OS/400 uses *Coded Character Set Identifiers* (CCSIDs) in addition to other system values. For example, OS/400 uses CCSIDs to convert accented characters like the French á and ê to English "a" and "e".

Most of the time, you can use your AS/400 with no concern at all about your system's CCSID. However, you may have to work with CCSIDs and conversion issues in situations like these:

- You have an AS/400 that supports more than one national language (e.g., English and Spanish on the same AS/400).
- You distribute software or files from your AS/400 to users in other countries.

And there's always that little problem of downloading spool files to the PC.

Code Pages and Code Points

The AS/400 ships with an extensive set of *Code Pages*. Code Pages are used to map characters in one character set to another. Code Pages are made up of *Code Points*, which are the hexadecimal values (x'00' to x'FF') that represent the characters.

Code Pages are located in library QUSRSYS. You can use the Work with Objects Using PDM command (WRKOBJPDM) to view the list of *TBL objects, which are the Code Pages.

System Values that Control Conversions

There are four system values that have a role in character set conversion. You access those values with the OS/400 Work with System Values command (WRKSYSVAL):

- QCCSID – coded character set identifier
- QCHRID – graphic character set and code page
- QLANGID – national language ID
- QCNTRYID – country ID

All of the values are set to defaults depending on the OS/400 language feature code that you install. For example, the U.S. English feature code is 2924, which sets QCHRID to code page 37, QLANGID to ENU, and QCNTRYID to US. (You can get an idea of where in the world the AS/400 is used by prompting for the QLANGID and QCNTRYID values in the WRKSYSVAL display.)

The QCCSID value is shipped with a value of 65535. This value is referenced by many other functions on the AS/400 where the CCSID is used (the reference is usually to *SYSVAL, which points back to QCCSID). Data encoded with the value 65535 cannot be translated from one code page to another.

Although IBM suggests that you can change your QCCSID System Value (for example, an appropriate value for U.S. English is 37), it does not seem to be necessary to change this. When you start a job on the AS/400, the job is assigned two attributes for CCSID:

- Job CCSID, taken from the user profile CCSID
- Default CCSID, based on the QCHRID value.

Unless changed, the user profile CCSID is based on the System Value QCCSID.

Files and CCSID

When you create a database file, the CCSID for the file is based on the Default CCSID for the job. If you have not changed anything on a U.S. English AS/400, the CCSID will be 37. The National Language Support manual mentions that database files created prior to V3R1 with CCSID 65535 are converted to the new CCSID when the release is installed. If you restore a file from a previous release to a V3R1 or higher AS/400, the file's CCSID is converted during the restore.

The Flat-file Exception

However, there is an exception for non-database files. An example is a file that you create to copy spool files into using the Copy Spool File command (CPYSPLF), with the intention of downloading the file to a PC.

The Help Panel for the CCSID parameter on the Create Physical File command (CRTPF) contains this description:

> This parameter is applicable only if no DDS is specified on the Source file prompt (SRCFILE parameter) and if a value is specified on the Record length prompt (RCDLEN parameter). If you specify a value other than the default value (*JOB) on this parameter, the default DDS shown on the SRCFILE parameter is not used, and a value must be specified on the RCDLEN parameter. Note: A file created with no DDS when FILETYPE(*DATA) is specified has a CCSID of 65535, regardless of the job CCSID value.

In other words, when you use CRTPF and specify a record length of 132 for the spool file, the file's CCSID is set to 65535 — and you can't change it. When you download that file to the PC, the CCSID value indicates that no conversion can take place for the file. Ultimately, you end up with hex data, unless you select the option to convert CCSID 65535 on the Client Access Data Transfer (in the File, Properties menu item).

References (IBM Manuals)

AS/400 Work Management, SC41-4306

AS/400 National Language Support, SC41-4101