

Progress Software

Report Title:	Work Request: W710010172	Page:	1
Requested By:	Paulo Lima (plima)	Printed:	10/18/07 08:35ET

W710010172 User: 61107519A DOUX FRANGOSUL SA
AGRO AVIC IN
Customer: Datasul S/A Request: 10//0/1/07 18:22
Caller: Roberta Favero Phone: [51] 47-38028345
Product: 0000106869 Query/RESULTS
Classification: Intlization Position: plima
Skill Group: am-database Territory: AMERICAS
Priority: 3-Medium Status: CUSTOMER
Resolved: By:
Codepage questions related to the Euro character in OE10

18-Oct-2007 08:33ET ACTION SOLUTION ENTRY: 20638 BY: Paulo Lima

Solution ID: 20638

TITLE: I18N. PROUTIL -C convchar charscan Functionality Explained

STATUS: Verified

GOAL:

How the Progress option PROUTIL -C convchar charscan functions in analyzing the database prior to conversion from existing character sets to new character sets that support Euro codepages.

GOAL:

PROUTIL -C convchar charscan Functionality Explained

GOAL:

PROUTIL -C convchar charscan

GOAL:

proutil <db> -C convchar charscan <codepage to convert to>

FIX:

Many new codepages are being created by governments, standards organizations, and hardware and operating system vendors to support the Euro.

Progress must support conversions from existing character sets to the new character sets. However, some of the new codepages cannot be converted to without dropping some of the characters that were in an existing codepage (for example, when the characters do not exist in the new codepage).

You must analyze your databases before you convert to new codepages to evaluate whether any data will be lost. To assist you with this analysis Progress provides new functionality (charscan) to the convchar analyze function. The convchar function presently scans the database to analyze the impact of doing a codepage conversion on the database.

If the new charscan option is specified and a list of character values provided, the characters are searched for throughout the database. You can then use the report of RECID's of records that contain any of these characters, to investigate how the records should be modified so the

Report Title: Work Request: W710010172
Requested By: Paulo Lima (plima)

Page: 2
Printed: 10/18/07 08:35ET

conversion does not lose data, or perhaps to decide not to convert the database.

The new option charscan is added to proutil -C convchar:

```
proutil <db> -C convchar charscan <codepage to convert to>  
<list-of-chars in db code page to report on>
```

In addition to the analysis that convchar analyze performs, the charscan option searches for the occurrence of any character from the provided list in every character field in the database, and reports the filename, fieldname and RECID of the match.

The report does not print the character that is matched. Once any of the characters is detected in a field, the search continues with the next field. Any list of the detected characters is misleading at this point because the remainder of the field is not searched and reported on. The total number of hits (fields with character matches) is printed at the conclusion of the analysis.

In Version 8.3, the charscan functionality applies to any database with a single-byte codepage. An error is provided if it is tried on a double-byte database. None of the proposed codepages for the Euro is a double-byte character set, so double-byte conversions are not an issue.

Beginning with Version 9.0A, the charscan functionality applies to any codepage, single-byte, double-byte and Unicode (UTF-8).

The character list is a quoted string of comma-separated numbers in either hex or decimal. These numbers represent character values in the database codepage. Hex values must begin with "0x". For example:

```
proutil <db> -C convchar charscan ibm850 "128,129,130"
```

or the same information provided in hex:

```
proutil <db> -C convchar charscan ibm850 "0x80, 0x81, 0x82"
```

The hex and decimal values can be mixed:

```
proutil <db> -C convchar charscan 1253 "128, 0xC2,0x7f, 122"
```

Legal values to scan for are in the range of 1-255 for single-byte codepages. The list of characters to search for can have as many as 10 values. If the charscan option is selected but there is no character list, the analyze function performs as in prior releases.

The search list is listed in the output, for example:

Report Title:	Work Request: W710010172	Page:	3
Requested By:	Paulo Lima (plima)	Printed:	10/18/07 08:35ET

```
proutil <db> -C convchar charscan ibm850 "128,129,130"
```

gives the following output (for an iso8859-1 db):

```
Charscan searching for iso8859-1 character: 128 0x80.  
Charscan searching for iso8859-1 character: 129 0x81.  
Charscan searching for iso8859-1 character: 130 0x82.
```

As an example, Microsoft codepage 1252 is identical to iso8859-1, except for a handful of characters. In particular, the Euro is assigned to value 128 in 1252 and does not exist at all in ISO8859-1. If you want to verify that a conversion from iso8859-1 to 1252 will not inadvertently cause an existing field with an iso8859-1 character with the value 128 to be misinterpreted as the Euro, you can issue the following:

```
proutil db -C convchar charscan 1252 "128"
```

If the database has any characters with the value 128 in its existing iso8859-1 codepage, the charscan report lists the records in messages such as:

```
Charscan found a character match in xxx.yyyy recid 999999.
```

or for a field with extents:

```
Charscan found a character match in xxx.yyyy[n] recid 999999.
```

New error messages for PROUTIL include:

```
Charscan error: Invalid <codepage> character: <dec> <hex>.
```

for example:

```
Charscan error: Invalid iso8859-2 character: 257 0x101.  
Charscan error: Invalid utf-8 character: 15708867  
0xefb2c3.
```

At the end, the total number of fields that had a match are listed:

```
Charscan match count: 99999
```

Errors will also indicate that the input list is too long, and (on Version 8.3A only) that a double-byte database cannot be scanned for characters.

NOTE:

References to Written Documentation:

Progress Software

Report Title: Work Request: W710010172
Requested By: Paulo Lima (plima)

Page: 4
Printed: 10/18/07 08:35ET

Progress Solutions:

20637, "The Euro (?) Currency and Progress"