Ruby - Feature #20959

Add a way to get codepage of console.

12/17/2024 03:42 PM - YO4 (Yoshinao Muramatsu)

Status:	Assigned
Priority:	Normal
Assignee:	windows
Target version:	3.5

Description

Abstract

Add a way to retrieve code pages of console.
On Windows, Encoding.find("locale") returns the console codepage.
To prepare for future changes, specify console instead of locale to get the encoding when a console code page is needed.

Background

On Windows, Encoding.find("locale") returns the console codepage.
This is different from locale in other environments. Also the name and content do not seem to match.
In the future, if we change the locale_encoding of the Windows port to the locale codepage of the C runtime library, we need to get the encoding for the console.
Strings received via pipe from cmd.exe or powershell are encoded in the console codepage.
This would be necessary when communicating with other programs via pipes.

Proposal

Make Encoding.find("console") return the encoding that represents the console codepage.

Background(continued)

Since Windows 10, UTF-8 support seems to be enhanced in the commandline environment.

- build 17134
  ucrt supports UTF-8 locale codepage setlocale(LC_CTYPE, ".utf8")
  Windows support UTF-8 ANSI Codepage(experimental) Beta: Use Unicode UTF-8 for worldwide language support
- build 18362
  support Set a process code page to UTF-8 via manifest.
- build 19041
  Time zone name holds in wchar_t internaly, so proper tzname can be obtained regardless of the Windows language setting.

This should not be all.
Through these, I feel that the Microsoft team recommends the use of utf-8. It also seems that Microsoft, which has previously depricated the ANSI version of the API, is treating it as a valid alternative to use in UTF-8.
If it becomes widespread on Windows for libraries to respect the C runtime library locale, ruby would do better to follow it.

Of course, it is good to be able to get the encoding of the console by specifying a console.

Future plan

By reducing differences from other platforms, bugs and extra code are hoped to be reduced.

- Do setlocale(LC_CTYPE, ".utf8"); in main.c (or refer LC_* environment variables).
- Encoding.find("locale") returns C runtime library locale.

Since strings obtained from the Windows system have a Unicode code range, the API to obtain fixed UTF-8 encoding remains unchanged.

Discussion

The code page we can get from Windows also has ACP and OEMCP, but are these necessary?
Is it reasonable to get locale_encoding if Encoding.find("console") is called on other platforms?

History

#1 - 12/25/2024 06:05 AM - nobu (Nobuyoshi Nakada)

- Assignee set to windows

- Target version set to 3.5

#2 - 01/07/2025 07:11 AM - hsbt (Hiroshi SHIBATA)

- Tags set to windows

#3 - 01/07/2025 07:14 AM - hsbt (Hiroshi SHIBATA)

- Tags changed from windows to win

#4 - 03/27/2025 01:15 AM - hsbt (Hiroshi SHIBATA)

- Status changed from Open to Assigned

#1 - 12/25/2024 06:05 AM - nobu (Nobuyoshi Nakada)

- Assignee set to windows