Technical Note: Encoding

Platform dependent Encoding Conventions

Version 1.01

PDF-TOOLS.COM
Premium PDF Technology

1 Overview

The encoding for files, command line and API outputs consider the platform depended conventions.

This technical note documents the concrete behavior of PDF Tools' software components.

2 Platforms

2.1 Windows

Output Files

Output files always use the UTF8 encoding and start with a byte order mark (BOM).

Command Line Output

Command line output always uses the system default Windows ANSI code page. This code page can be different on different computers.

2.2 Unix

File and Command Line Tools

Both files and command line output use the application's current locale[1] settings and no byte order mark (BOM). The application's locale setting should be set using the `setlocale` function http://www.opengroup.org/onlinepubs/007904875/functions/setlocale.html).

The default is the standard C locale as specified in the ANSI C standard.

Command Line Tools

The command line tools set the locale to the value specified by the system's environment variables

C/C++ APIs

The C/C++ APIs behave like the ISO C standard library and sets no locale. It is the application developer's responsibility to choose an appropriate locale.

JAVA APIs

The Java runtime environment sets the locale.

---

[1] A set of parameters that define the regional settings of the user.

3 Contact

PDF Tools AG
Kasernenstrasse 1
8184 Bachenbülach
Switzerland
http://www.pdf-tools.com
pdfsales@pdf-tools.com