

Linux Encoding Renamer

Over the course of the past few years I have moved between Linux and Windows file systems. I can't recall exactly in what sequence but it is of little consequence. Let us not forget that not only operating and files systems have been changed but also the languages natively supported by said systems. So most of my files with non ASCII characters have got weird names. This has mainly been a problem for my music files which are both numerous and frequently expressed in non ASCII characters. 150 files were eventually renamed and a similar number were deleted out of expediency(I can just rip them again).

I added some functionality to an existing Java program I wrote for similar administrative tasks and painstakingly mapped incorrect characters to their proper counterparts. I provide you with the table on the right to give you an idea why character encodings are important and also a source of premature death in computer professionals. Note how the same letter can be corrupted in more than one way and that one letter is not replaced but simply given a nice prefix character. I thought the recurring number 9500 implied UTF had been involved at some point but then I remembered that Java uses UTF encoding internally, so that must be why those numbers show up.

┆ñ	ä
ΓÇPтs	ä
┆л	ü
┆ü	ü
┆ll	ö
┆û	ö
┆â┆ll	ö
ΓÇ¥	ö
┆Ñ	å
┆â┆Ñ	å
┆l	'
┆r	é
┆f	ß
┆í	ß
┆l	?
┆z	è
┆á	à
┆°	Ç