TextAutoDecode

# Objectives

Read a text file, automatically detecting text encoding

Returns Result<TextData, io::Error>

struct TextData {

text: Option<String>,

encoding: TextFileEncoding,

}

enum TextFileEncoding {

NotText, // Binary or unrecognized text (chars in 0..31 other than \r \n \t)

Empty, // File is empty

ASCII, // Only 7-bit characters

EightBit, // ANSI/Windows 1525 or other

UTF8,

UTF8BOM, // Starts with EF BB BF

UTF16, // No BOM, assume UTF16LE (should use heuristics)

UTF16LE, // Starts with FF FE (Windows)

UTF16BE, // Starts with FE FF

}

- Read first 1000 bytes max to detect text encoding, if it's not recognized as text, return quickly

- Otherwise read and decode full text (without BOM), return text and TextFileEncoding

Crate can also expose a function to detect EOL from a &str, returning an enum:

enum TextFileEOLStyle {

None,

Windows,

Unix,

Mac,

Mixed,

}