

fountain-parser v.0.1.0.0 README

Synopsis

`fountain-parser` is a small parser library for the [FOUNTAIN](#) screenplay format, fully supporting 1.1 version [syntax](#) and producing a simple, easy to grok AST.

`fountain-parser` is written in [HASKELL](#) and it uses the [MEGAPARSEC](#) library for parsing.

Disclaimer

Currently, this is *pre-alpha* software, not yet usable in productive form.

This software is distributed *as-is* under the terms of the BSD THREE-CLAUSE LICENSE. See the [LICENSE](#) file for more details.

Motivation

The [Developers section](#) of the Fountain site provides a link to a [parsing library](#) in OBJECTIVE C. This presents a portability issue: there *are* projects that make it possible to bridge OBJECTIVE C and HASKELL, they're platform- or framework-specific. That library informs this project in matching the different Fountain entities even as it uses different parsing methods.

Prospective Related Projects

`fountain-parser` aims to power a series of command-line utilities for conversion from FOUNTAIN to a series of convenient formats (like `.TEX`) without intervention from thirds.

My software already supports Fountain

The [Apps section](#) of the FOUNTAIN site lists software that also imports or exports the format. There's a caveat: most are either *cloud-based* and/or *proprietary*. By favoring (mostly) open formats, *fountain-parse* allows integration into many FLOSS tools, enabling entirely non-proprietary workflows and helping the creation of compound documents such as production bibles.

Implementation Specifics

- As per the syntax guide:
 - This library expects Fountain text to be encoded in UTF-8.
 - Tabs are converted into **four** spaces.
 - Your line-spacing is respected.

- Initial spaces are ignored everywhere except in action lines.
- A line with two spaces doesn't count as an empty line.
- All parsing functions expect `Text` inputs. File I/O is left to the application or framework.
- Varying-width `UNICODE` spaces are either converted into regular spaces or suppressed if they're hairline- or zero-width.
- Vertical tabs and form-feed characters are interpreted as line changes. For vertical spacing, use multiple blank lines and/or the Fountain form feed character sequence instead.
- The parser keeps everything: notes, boneyards, sections and synopses. Some possible conversion targets have equivalents to those, thus it might be desirable to preserve them.

Tentative Formal Grammar

The following is an attempt to formalize the syntax in ABNF, drawing from the syntax guide and OBJECTIVE C implementation. It incorporates `UNICODE` codepoints and it tries to err in the side of lenience.

; The grammar is currently very ambiguous, requiring unrestricted lookahead or backtracking.

fountain-screenplay = *empty-line (script-content / cover-page 1*empty-line script-content)

cover-page = 1*cover-entry

cover-entry = cover-key *space ":" *space cover-value

cover-key = "TITLE" / "CREDIT" / "AUTHOR" / "SOURCE" / "DRAFT DATE" / "CONTACT"

cover-value = single-value / multi-value

single-value = 1*non-newline newline

multi-value = newline 1*indented-sub-value

indented-sub-value = (htab / 2*8(space)) 1*non-newline newline

script-content = *(section-indicator / master-scene / synopsis)

section-indicator = 1*"#" *space 1*non-newline newline empty-line

master-scene = master-scene-heading scene-content

master-scene-heading = int-ext scene-description *scene-number *space newline empty-line

int-ext = ("I" (["."] "/" ["E"] / ("NT" (["."] "/EXT")) / "E" ("ST" / "XT")) (["."] / space)

scene-description = *space 1*non-newline-or-hash

scene-number = "#" 1*scene-number-character "#" *space

scene-number-character = alphanumeric / "-" / "."

uppercase-letter = %x41-5A

digit = "0" / "1" / "2" / "3" / "4" / "5" / "6" / "7" / "8" / "9"

alphanumeric = alpha / digit

```

space = " "
/ %x00A0      ; non-breaking
/ %x2000-2009 ; varying-width Em/En-based spaces
/ %x202F      ; narrow non-breaking
/ %x205F      ; mathematical middle-space
/ %x3000      ; Ideographic space
; These are turned into the regular space; we're trying to imitate a typewriter.
; Hairline or zero-width spaces and joiners are pre-filtered out previous to parsing.
; Same goes for control characters.

newline = cr [lf]
/ lf [cr]
/ %x0B      ; We interpret vertical tabbing as a newline too
/ %x0C      ; As well as form-feeding
/ %x0085    ; Unicode next-line
/ %x2028    ; Unicode line-separator
/ %x2029    ; Unicode paragraph-separator
; These are all converted into the POSIX newline "\n".

empty-line = *space newline

character = <any Unicode character>

non-newline = <character - newline>

non-newline-or-hash = <non-newline - "#">

```

Building

GHC 9.6.7 and CABAL 3.0 (or greater) are required to compile and run the test suite (once implemented.)

The project uses the GHC2021 language default. While it might be possible to compile it in earlier versions than 9.6.7, this default is only available since 9.2.1., so that constitutes a hard version limit for those who want to try other versions.

Some of the included scripts require `make`, `awk` and other similar utilities usually found in LINUX or LINUX-like environments (e.g., [MSYS2](#).) However, nothing prevents the user from running `cabal`, `pandoc` or `pdflatex` as shown in the [Makefile](#).

Contact

Please [create an issue](#) if you find a bug.

I can be reached directly at *10951848+CübQfJúdãhsLîòn ä(t) users/noreply/g̃ithub/còm* (without accents and replacing slashes by periods.)