Adam Laschkewitsch
Mechanical Engineer

Regex-Based Text Correction for EPUB Files

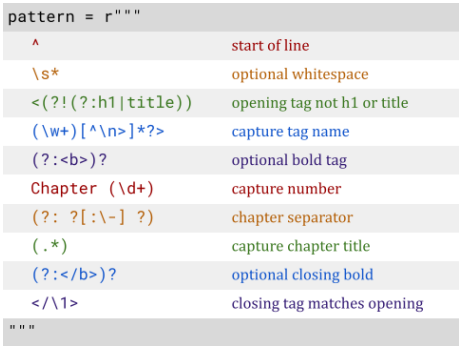
Adam Laschkewitsch — Regex-Based Text Editor | Mar 2024 – Present

Overview

Used regular expressions and lightweight Python scripting to **automate grammar corrections and unify text formatting** across EPUB files. Custom regex patterns and targeted rules streamlined cleanup of large text collections, significantly reducing manual effort. The project reinforced the value of scripting for automation and deepened understanding of pattern design through practical, self-taught application.

Tools & Skills

Regular Expressions (Regex), Basic Python Scripting, Calibre EPUB Editor, HTML, Text Pattern Recognition, Automation of Repetitive Tasks, Self-Directed Learning



pattern = r"""	
^	start of line
\s*	optional whitespace
<(?!(?:h1|title))	opening tag not h1 or title
(\w+)[^\n>]*?>	capture tag name
(?:)?	optional bold tag
Chapter (\d+)	capture number
(?: ?[:\-] ?)	chapter separator
(.*)	capture chapter title
(?:)?	optional closing bold
</\1>	closing tag matches opening
"""	

Pattern breakdown for chapter title cleanup.

Objective

Apply regular expressions and lightweight Python scripting to automate bulk grammar corrections and standardize text formatting across EPUB files.

Highlights

- Automated bulk editing of large text collections, reducing manual correction time significantly.
- Custom regex patterns designed to correct common grammar mistakes and unify inconsistent formatting.
- Targeted formatting rules for dialogue, emphasis, and spacing tailored to personal reading preferences.
- Introduced **lightweight Python scripts** to streamline multi-file processing and ensure consistent application.
- Self-taught regex fundamentals, allowing fast iteration on complex pattern-matching problems.

Reflections

Learning regex through **hands-on application** helped solidify abstract concepts and **revealed edge cases** rarely encountered in simpler examples. Automating text cleanup highlighted the **value of small-scale scripting for repetitive tasks** and motivated further exploration of Python. Designing effective patterns required **balancing precision and generality**; too broad and they risked altering valid content, too narrow and they failed to catch real issues.

alaschke.com
linkedin.com/in/alaschke
alaschke+portfolio@outlook.com