Practical Python— Regular Expressions

In this project, you will use **Python and Regular Expressions (RegEx)** to explore and analyze books from **Project Gutenberg**, a free library of public domain texts.

You will learn how to:

1. Extract **numbers, dates, and years** from a book
2. Find **character names, places, and full names**
3. Identify **dialogues, quotes, and sentences containing specific words**
4. Detect **punctuation, hyphenated words, and capitalized acronyms**
5. Locate **URLs, email-like patterns, or other special text patterns**

The project will guide you through at least **15–20 different RegEx patterns**, helping you practice **real-world text analysis**.

Exercises

Exercise 1 — Download and Read a Book

- Select a book from Project Gutenberg (plain text version).
- Download the .txt file and read it using Python.

Exercise 2 — Basic Information Extraction

- Extract **years** (4-digit numbers).
- Extract **all numbers**.
- Extract **capitalized words** (potential names and places).
- Extract **full names** (two consecutive capitalized words).

Exercise 3 — Dialogue and Sentence Extraction

- Extract all **dialogues** (text inside quotes).
- Extract **chapter headings**.
- Extract **sentences containing a specific word**.

Exercise 4 — Advanced Pattern Extraction

- Extract words **ending with "ing"**.
- Extract **punctuation marks**.
- Extract **words in all caps**.
- Extract **italicized words** (if the book uses *word* for emphasis).
- Extract **words starting with vowels**.
- Extract **hyphenated words**.
- Extract **decimal numbers**.
- Extract **short words (1–2 letters)** and **long words (8+ letters)**.

- Extract **URLs or email-like patterns** (if present).

Advanced Exercise (Optional) — Build a Custom RegEx Search Tool

- Create a Python CLI tool that:
    - Accepts a book file as input.
    - Accepts a user-defined RegEx pattern.
    - Displays the number of matches and the first 20 results.