My Grep Implementation

Overview

This project implements a simplified version of the Unix grep command in C, called my_grep. The program searches for a specified pattern in an input file or standard input and outputs the matching lines with various options for customization.

Features

Basic Features

- **Pattern Matching**: Searches for fixed strings in the input.
- **Options**:
    - -A NUM: Prints NUM lines after the matching line.
    - -b: Displays the byte offset of each matching line.
    - -c: Prints only the count of matching lines.
    - -i: Ignores case when matching.
    - -n: Prints line numbers alongside the output.
    - -v: Inverts the match to print lines that do not contain the pattern.
    - -x: Matches only whole lines that exactly match the pattern.

Regular Expression Features

- **Enhanced Pattern Matching**: When using the -E option:
    - Patterns must be enclosed in quotes.
    - Supports alternation (e.g., str1|str2).
    - Matches any character with . (dot), but dots cannot appear within parentheses.
    - Supports range matching with [x-y], where both x and y must be in the same category (e.g., both uppercase letters, lowercase letters, or digits).
    - No nested parentheses or brackets within other brackets or parentheses.

Requirements

The program supports:

- Regular strings and basic regular expressions.
- Input from standard input (stdin) or a specified file.
- Proper handling of special characters when escaping.

Installation

To compile the program, use the provided Makefile, simply run: "make"

This will generate the executable file my_grep.

Usage

To run the program, use the following syntax:

./my_grep <OPTIONS> <PATTERN> <FILE>

Example:

Execution of the line: my_grep –n –i -E ‘o\.pdf’ 2024.txt

-n: indicates that line numbers should be printed along with matching lines.

-i: means that the search should be case-insensitive.

-E: indicates that extended regular expressions are being used.

'o\.pdf': The search pattern. Here, it is looking for the string "o.pdf". The backslash before the dot (\.) is used to escape the dot, making it a literal dot rather than a wildcard character.

2024.txt: The input file being searched.

If 2024.txt contains:

1: This document is o.pdf.

2: Here is a link to O.PDF.

3: Another file is test.pdf.

4: Look at this: o.PDF file.

The output is expected to be:

1: This document is o.pdf.

2: Here is a link to O.PDF.

4: Look at this: o.PDF file.

Notes

- All functions are implemented without relying on standard library functions for regular expressions.
- The code is formatted and linted using clang-format and clang-tidy to maintain quality.