MultiGrep

By David Hiatt
(dhiatt@pdx.edu)

GNU General Public License

Audience

- Command line tool for programmers
-

- Reads in a file of patterns followed by their replacements.
- Searches for all patterns in any number of files and replaces them with their specific replacement.

Description

- Creates a prefix tree (trie) composed of elements from every pattern.
- Uses a producer and consumer thread synchronization model.
- Performs a single comparison per byte per file.

Alternatives

- Sed, awk, bash, perl, etc
- IDE refactoring
- Manually editing
- Possibly many other methods

Motivation

- Client localization for an open source online game.
- Required replacing 73,829 unique strings found any number of times in a large code base.

Lessons

- Project works very well when using deterministic patterns.
- Regular expressions introduces potential ambiguity.

Future

- Add more support for regular expressions
- Convert to C using pthreads for faster performance
-
-
-

Contact

- https://github.com/D-Hiatt/CS461
- dhiatt@pdx.edu