Constant-Delay Enumeration for Nondeterministic Document Spanners

 63 commits   **3** branches   **0** releases   **2** contributors

master ▾   New pull request   Find File   Clone or download ▾

a3nm typo		Latest commit f97d174 Aug 29, 2019
📁 benchmarks	add benchmarks to track performances over next versions	May 14, 2019
📁 src	fix output offset when utf8 data is part of the text	Jul 26, 2019
📄 .gitignore	tests: split match tests and enumeration tests	May 14, 2019
📄 .rustfmt.toml	add rustfmt config	Jul 22, 2019
📄 Cargo.lock	add naive algorithms enumerating matching subwords	Jul 23, 2019
📄 Cargo.toml	add naive algorithms enumerating matching subwords	Jul 23, 2019
📄 README.md	typo	Aug 29, 2019

📖 **README.md**

Constant-Delay Enumeration for Nondeterministic Document Spanners

This tool allows you to find efficiently all matches of a regular expression in a string, i.e., find all contiguous substrings of the string that satisfy the regular expression (including overlapping substrings).

The tool is being actively developed and has not been thoroughly tested yet. Use at your own risk.

It is the reimplementation of the previous Python prototype.

Requirements

It has been tested and developed using `rustc 1.34` and `cargo 1.34`. It will not work with older Rust versions shipped by some Linux distributions, e.g., with version 1.32. You can check your rust version with `rustc --version`, and install manually a more recent Rust version from the Rust website.

Specific library requirements can be found in *Cargo.toml* and *Cargo.lock*.

Usage

The quickest way is to run the program through Cargo.

```
# Display all occurences of a pattern (regexp) in a file
cargo run --release -- [regexp] [file]
```