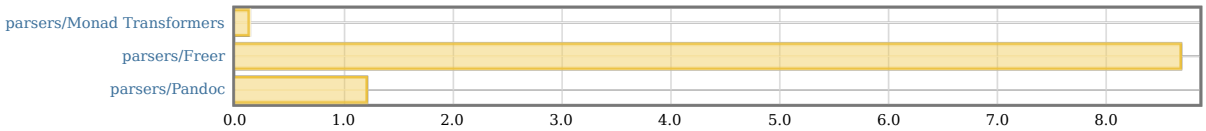
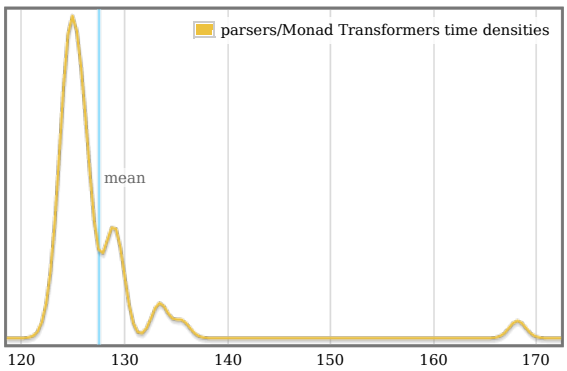
Markdown parsers performance measurements

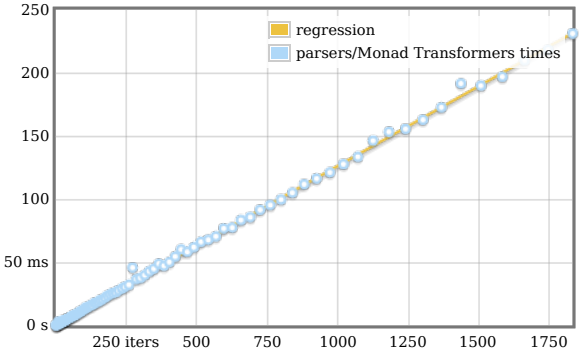
overview

The benchmarking was done on a Markdown file of size 355 bytes.



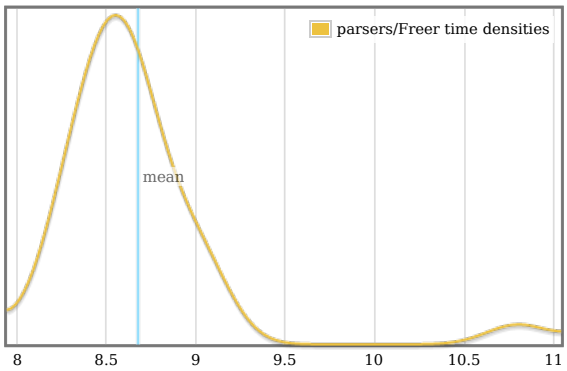
parsers/Monad Transformers

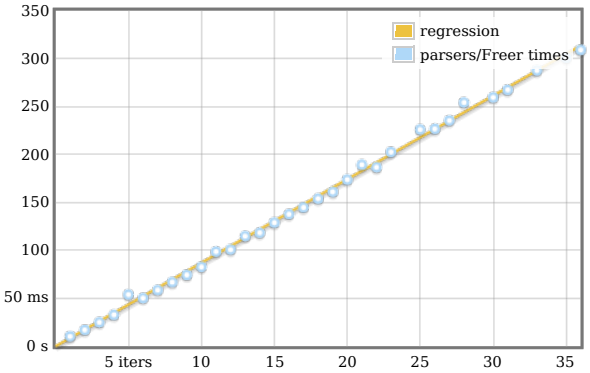




	lower bound	estimate	upper bound
OLS regression	125 µs	126 µs	128 µs
R² goodness-of-fit	0.998	0.999	1.000
Mean execution time	126 µs	128 µs	131 µs
Standard deviation	2.58 µs	6.95 µs	13.9 µs

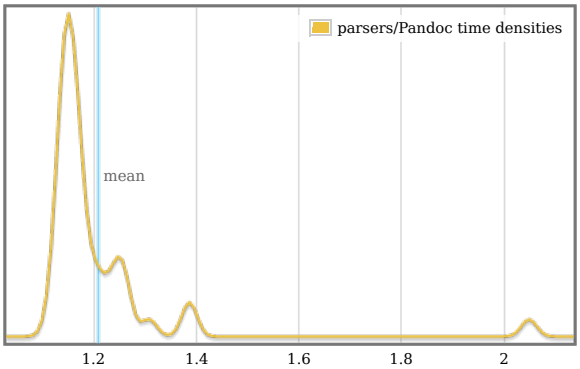
parsers/Freer

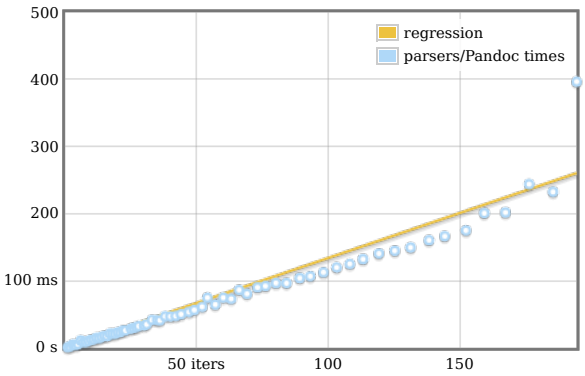




	lower bound	estimate	upper bound
OLS regression	8.58 ms	8.71 ms	8.90 ms
R² goodness-of-fit	0.996	0.998	0.999
Mean execution time	8.56 ms	8.68 ms	9.06 ms
Standard deviation	209 µs	466 µs	877 µs

parsers/Pandoc





	lower bound	estimate	upper bound
OLS regression	1.18 ms	1.34 ms	1.59 ms
R² goodness-of-fit	0.889	0.928	0.997
Mean execution time	1.18 ms	1.21 ms	1.30 ms
Standard deviation	57.1 µs	148 µs	291 µs

understanding this report

In this report, each function benchmarked by criterion is assigned a section of its own.

- The chart on the left is a kernel density estimate (also known as a KDE) of time measurements. This graphs the probability of any given time measurement occurring. A spike indicates that a measurement of a particular time occurred; its height indicates how often that measurement was repeated.
- The chart on the right is the raw data from which the kernel density estimate is built. The *x* axis indicates the number of loop iterations, while the *y* axis shows measured execution time for the given number of loop iterations. The line behind the values is the linear regression prediction of execution time for a given number of iterations. Ideally, all measurements will be on (or very near) this line.

Under the charts is a small table. The first two rows are the results of a linear regression run on the measurements displayed in the right-hand chart.

- *OLS regression* indicates the time estimated for a single loop iteration using an ordinary least-squares regression model. This number is more accurate than the *mean* estimate below it, as it more effectively eliminates measurement overhead and other constant factors.
- *R² goodness-of-fit* is a measure of how accurately the linear regression model fits the observed measurements. If the measurements are not too noisy, R² should lie between 0.99 and 1, indicating an excellent fit. If the number is below 0.99, something is confounding the accuracy of the linear model.
- *Mean execution time* and *standard deviation* are statistics calculated from execution time divided by number of iterations.

We use a statistical technique called the bootstrap to provide confidence intervals on our estimates. The bootstrap-derived upper and lower bounds on estimates let you see how accurate we believe those estimates to be. (Hover the mouse over the table headers to see the confidence levels.)

A noisy benchmarking environment can cause some or many measurements to fall far from the mean. These outlying measurements can have a significant inflationary effect on the estimate of the standard deviation. We calculate and display an estimate of the extent to which the standard deviation has been inflated by outliers.

colophon

This report was created using the criterion benchmark execution and performance analysis tool.

Criterion is developed and maintained by Bryan O'Sullivan.