Experiments In Parallelisation In R

Peter von Rohr

2018-05-03

Disclaimer

This document describes a first set of experiments with parallelisation in R. We are using pair-wise comparisons between strings using the R-package `stringdist`[1]. There is a paper on stringdist[2]

[1] GitHub `https://github.com/markvanderloo/stringdist`

[2] R-Journal `https://journal.r-project.org/archive/2014-1/loo.pdf`

Introduction

There are two motivations behind the experiments shown in here

1. string comparisons is a hard and an important problem
2. when it comes to increasing the preformance above a certain level, there is no way around parallelisation.

Background

String comparisons

The background and the importance of string comparison is described in [van der Loo, 2014]. Furthermore, the cited paper also describes a list of R-packages in the context of string comparison, are given.

Parallelisation

The technology of parallelisation used in `stringdist` is multithreading based on openmp[3]

[3] On line course taught by Tom Mattson from Intel, `https://www.youtube.com/playlist?list=PLLX-Q6B8xqZ8n8bwjGdzBJ25X2utwnoEG`

Experiments

This section describes concrete applications of the `stringdist` functionality to different types of data.

References

M. P.J. van der Loo. The stringdist package for approximate string matching. *R Journal*, 2014. URL `https://journal.r-project.org/archive/2014-1/loo.pdf`.