



Onni Koskinen, Arturs Polis, and Lari Rasku

We have implemented a family of three different string sorting algorithms in Python and compared their performance against Python's native Timsort using a variety of different datasets.

MSD RADIX SORT

Our implementation uses a fixed alphabet size of 256 and falls back to ternary quicksort when the size of the bucket drops below it.

QUICKSORT ALGORITHMS

Quicksort text Quicksort text Quicksort text Quick-
sort text Quicksort text Quicksort text Quicksort
text Quicksort text Quicksort text Quicksort text
Quicksort text Quicksort text Quicksort text Quick-
sort text Quicksort text Quicksort text

BURST SORT

Burst sort text Burst sort text Burst sort text
Burst sort text Burst sort text Burst sort text Burst
sort text Burst sort text
Burst sort text Burst sort text
Burst sort text Burst sort text
Burst sort text

TEST DATA

A 100MB and a 200MB sample of each dataset was used. The ENGLISH datasets were not used as-is, but with each word split on its own line, in order to make the algorithms sort individual words and not entire lines. The statistics file documents some stringological properties of these

ref1 <https://sites.google.com/site/ranjansinha/home>
ref2 <http://www.cs.mu.oz.au/~rsinha/resources/data/s>

REFERENCES

- [1] U. Lauther and T. Lukovszki. Space efficient algorithms for the Burrows-Wheeler backtransformation. In *Proc. 13th Annual European Symposium on Algorithms*, volume 3669 of *LNCS*, pages 293–304. Springer, 2005.
 - [2] J. Seward. Space-time tradeoffs in the inverse B-W transform. In *Proc. IEEE Data Compression Conference*, pages 439–448. IEEE, 2001.
 - [3] U. Lauther and T. Lukovszki. Space efficient algorithms for the Burrows-Wheeler backtransformation. In *Proc. 13th Annual European Symposium on Algorithms*, volume 3669 of *LNCS*, pages 293–304. Springer, 2005.
 - [4] J. Seward. Space-time tradeoffs in the inverse B-W transform. In *Proc. IEEE Data Compression Conference*, pages 439–448. IEEE, 2001.
 - [5] U. Lauther and T. Lukovszki. Space efficient algorithms for the Burrows-Wheeler backtransformation. In *Proc. 13th Annual European Symposium on Algorithms*, volume 3669 of *LNCS*, pages 293–304. Springer, 2005.
 - [6] J. Seward. Space-time tradeoffs in the inverse B-W transform. In *Proc. IEEE Data Compression Conference*, pages 439–448. IEEE, 2001.