Algorithms in Computational Biology and Sequence Analysis

Module 2: Strings and exact matching

Demo - Suffix array construction

Dr. Chirag Jain
Department of Computational and Data Sciences
Indian Institute of Science
Web: https://at-cg.github.io/
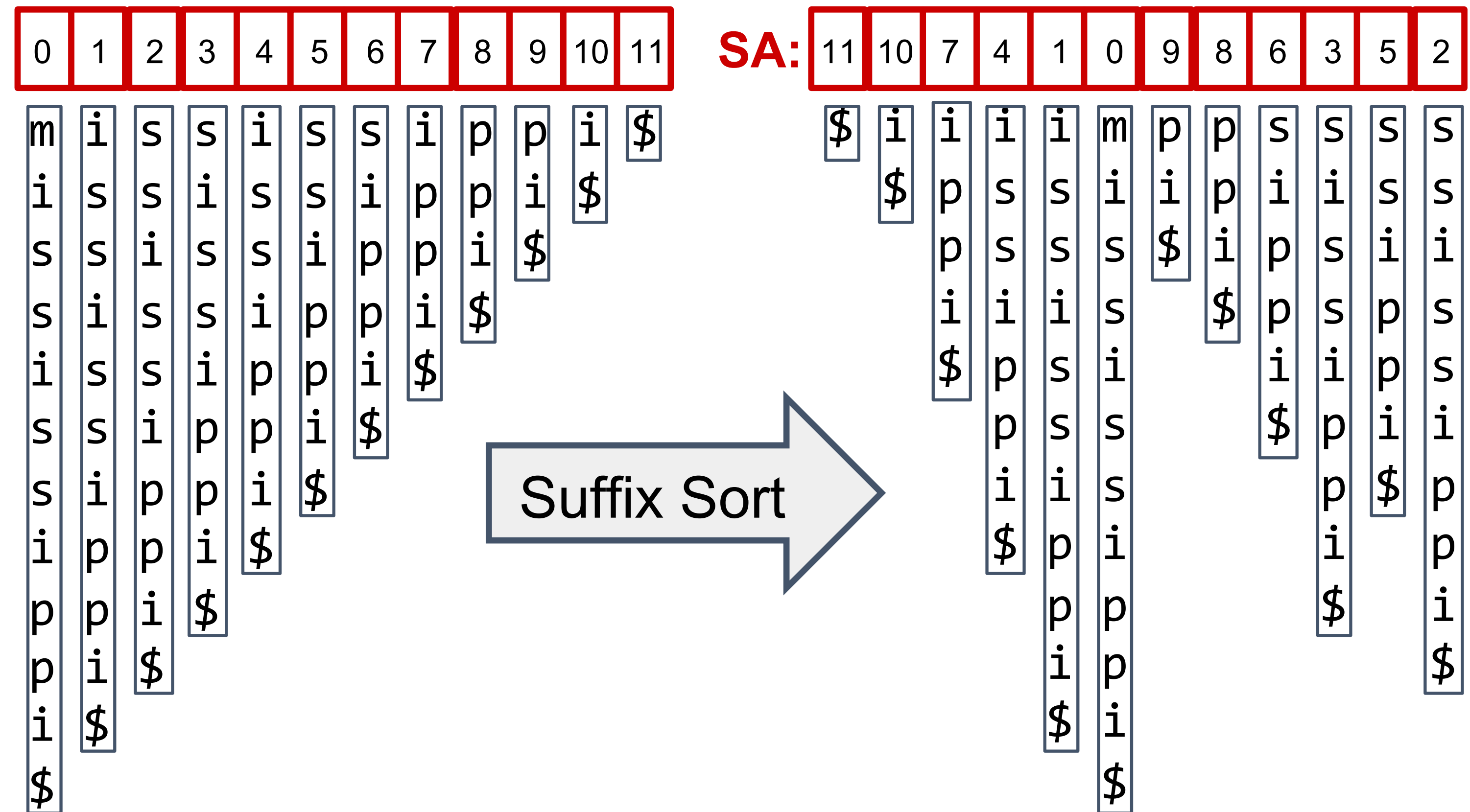
Teaching Assistant:

Daanish Mahajan
Interdisciplinary Mathematical Sciences
Indian Institute of Science

Recap

$SA(T)$: Lexicographically ordered array of all suffixes of text T

$T = \texttt{mississippi\$}$

Suffix Sort

Agenda
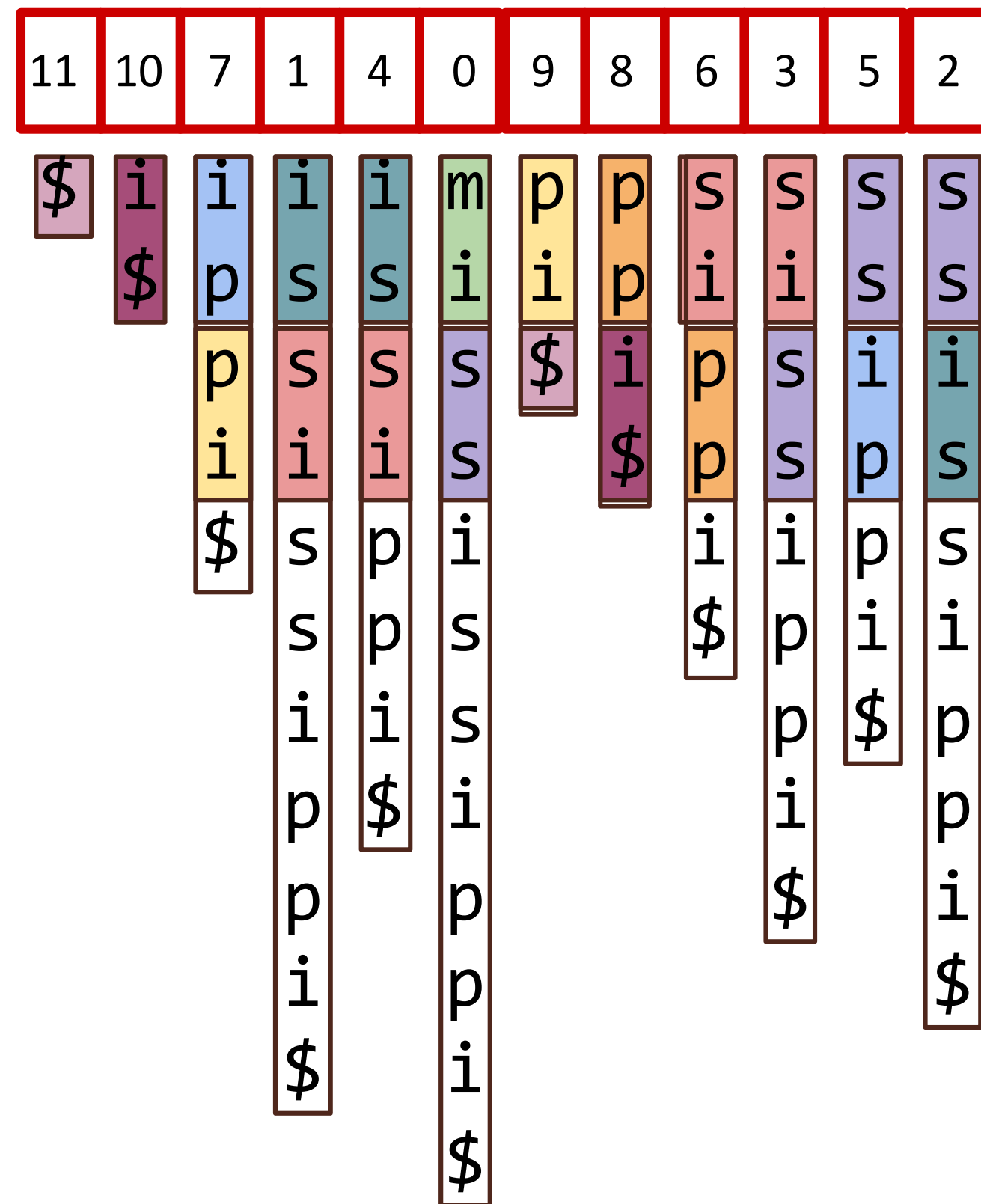
Coding algorithms to construct $SA(T)$

Naive Sorting

$O(|T|^2 \log |T|)$

Prefix doubling algorithm with naive sorting

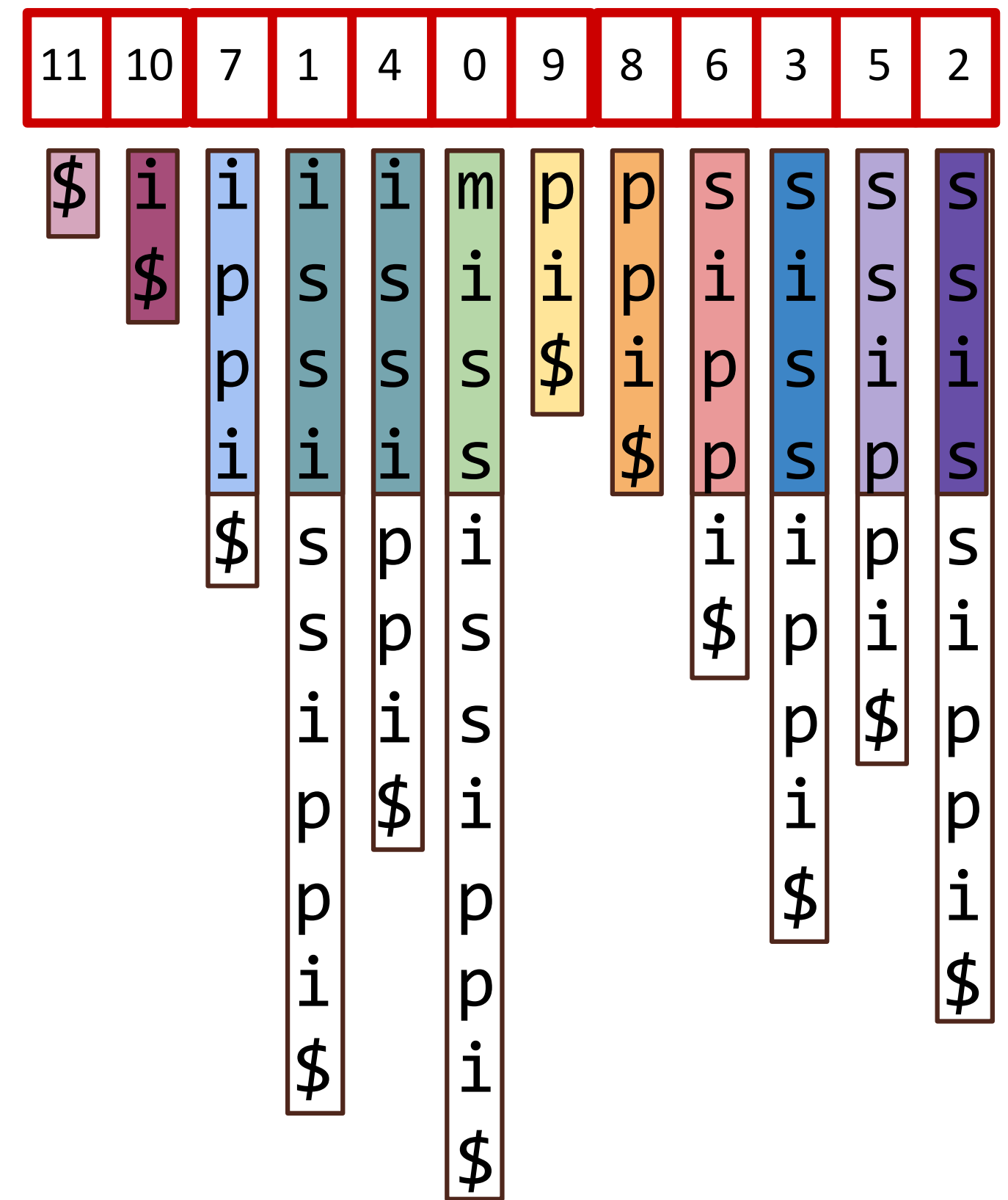$O(|T| (\log |T|)^2)$

Prefix-doubling

Sort strings by ID pairs

Now suffixes are sorted by prefixes of size 4, recalculate bucket IDs

Update the "inverse array"