Suffix Tree

$O(n^2)$ time & space $\leftarrow$ naive

$O(n)$ space & time $\leftarrow$ Ukkonen's algo

$\Sigma = \{A, B\}$


$\text{2nd suffix: } [2, e]$
2) Index matching:

- $|q| = l$ ← length of the query
- $d ← d + 1$ mismatches
- $d < e$

**Naive Approach:**

$q = BA$

$d = 1$

use the

Sphinx tree to

spell our the

neighbor "with

reserve count"
2) Genome Alignment

MUM

\[ d = 3 \]

\[ q \text{ must have an exact match of length } \left\lfloor \frac{d}{d+1} \right\rfloor = w \]

Generate all quene \( q \) len \( w \)

\[ \rightarrow \text{ look for an exact match} \]

\[ \rightarrow \text{ extend to longer matches} \]

[Diagram showing alignments with lengths 3x10^9 and 2.5x10^9]

\[ O(n) \text{ time!} \]
Maximal $\Rightarrow$ it cannot be extended to left or right in either string

Unique $\Rightarrow$ one occurrence in each string

1. $12345^T$
2. $ABAAB$ $ABAAB$

$N: \quad \overbrace{BAABABA}^{12345}$

1) Build ST for one of the strings
2) Add $2^\text{nd}$ string to the suffix tree

Unique matches: $ABAAB$ $(3, 5)$ $\checkmark$

$ABAAB$ $(1, 4)$ $\checkmark$
1) find the MUMS:

\[
\frac{O(n) + O(m)}{\text{suffix tree}} \quad \text{maximal clean}
\]

\[
O(\sqrt{n} |\Sigma|) \quad \text{real loss}
\]

over line \# q

unique matches (m)

2) find the maximum weighted subsequence of segments between lines 2 and same chain DP

\[\text{suffix away s}\]

sorted linear suffix of a DB string

\[1 2 3 4 5 6 7 8 9\]

\[A B A B A B A B \]$\]

\[\text{total order on } \Sigma u\]

$< A < B$
q: $O(\log n)$

$\frac{O(1) \log n}{?}$

DB size + $O(\# \text{matches})$

$O(\log n)$

How to do this?