

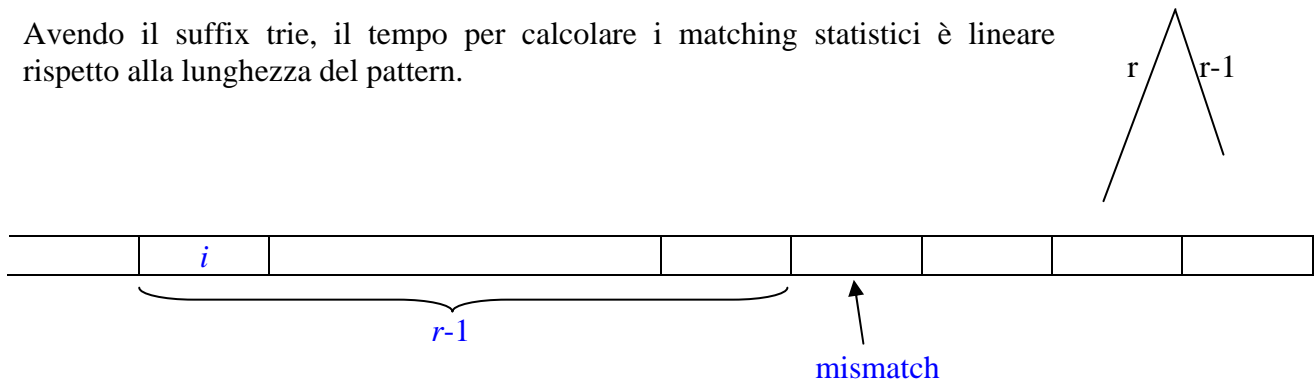
CALCOLO DEI MATCHING STATISTICI tra un pattern P fissato e il testo T , con $|T|=m$

NEI SUFFIX TRIE

Si sfruttano i suffix link, che nel trie esistono sempre, in modo tale da non rileggere i caratteri del testo per i quali si è già avuto il match: al più saranno riletti i caratteri che non sono nel pattern. Appena si verifica un mismatch pattern / testo sul nodo v alla posizione r , durante il calcolo di $ms(i)$,

- si aggiorna l'array dei matching statistici in posizione i ,
- si segue il suffix link $s(v)$ del nodo v ,
- si continua il match da quel carattere in poi senza rileggere gli $r-1$ caratteri già confrontati

Avendo il suffix trie, il tempo per calcolare i matching statistici è lineare rispetto alla lunghezza del pattern.



NEI SUFFIX TREE

Per il calcolo dei matching statistici utilizzando un suffix tree \mathcal{T} per il pattern P ed una copia del testo T , si utilizza il trucco dello “*skip / count*”. Se durante la fase di match ci si ferma in mezzo ad un arco,

- si risale al nodo precedente contando i caratteri (n) fino ad esso,
- si segue il suffix link $s(v)$ di v ,
- si saltano da $s(v)$ i primi n caratteri.

Non è necessario riconfrontare tali n caratteri perché tale confronto è già stato fatto precedentemente, prima del mismatch. Inoltre, è immediato osservare che gli n caratteri al di sotto del nodo v sono uguali agli n caratteri al di sotto di $s(v)$, proprio per l'esistenza del suffix link.