1 DICTIONARY MATCHING

A dictionary is a set of finite strings, $X = \{x_0, x_1, ...x_{k-1}\}$. Matching involves locating occurrences of X in any text y.

1.1 TRIE

A trie $\mathcal{T}(X)$ is a digital tree whose branches are labelled by strings of X and is the basis of the dictionary matching automaton (DMA). The nodes of the tree are prefixes of strings in X.

```
1 TrieState[] Children                          // children trie states
  char Incomming                          // incoming letter to this state
  TrieState Failure                                // one failure node
  bool IsTerminal                        // is this state terminal
```

Figure 1: TrieState — Trie State Data Structure

```
1 if p.Transitions[c] exists then
2 |   return q                                   // child exists return it
3 end
4 if p.Failure exists then
5 |   return TrieNextState (p.Failure,a)     // recursively follow failure link
6 end
7 return root of trie                                          // zero state
```

Figure 2: TrieNextState(p.Failure,a) Compute Next State Trie

```
1 q[i] ← a state representing string s[i]
2 q[j] ← a state representing string s[j]
3 if (s[j] longest proper suffix of s[i]) || (s[i] prefix of x ∈ X) then
4 |   q[i] points to the state q[j]
5 end
```

Figure 3: Compute failure links algorithm;

Failure links are computed top down, and then each node at given level. For a given node s_j with an incoming edge of letter c the failure links are computed as:

1. Go to s_j parents' failure.
2. If an outgoing edge of letter c is present follow and place failure link to here.

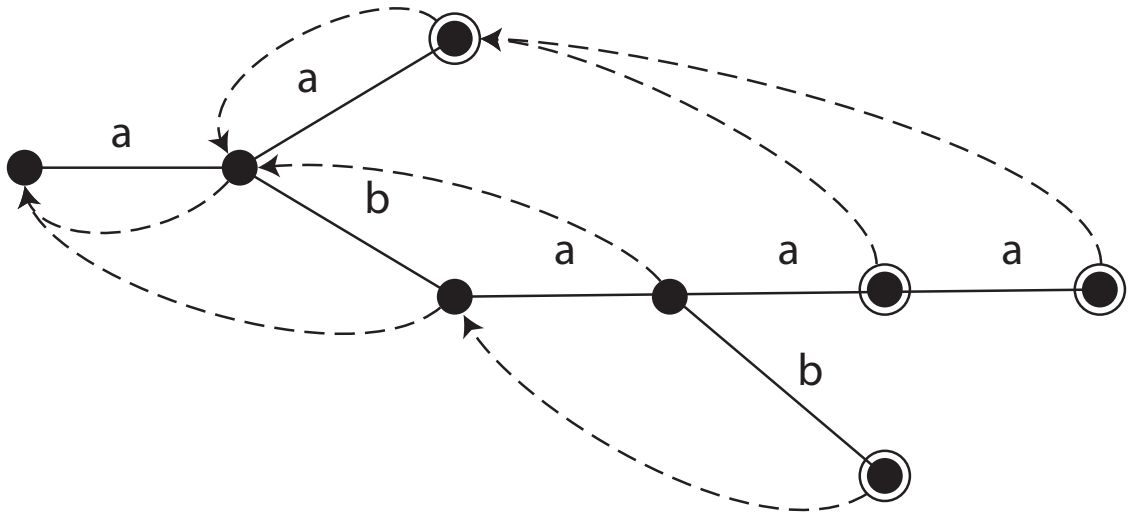

Figure 4: $\mathcal{T}(\{aa, abaaa, abab\})$

3. If an outgoing edge of letter c is not present repeat this process using current state.

The optimized failure links are computed from a DMA with the failure links already computed.

- if (current sate outgoing edges = failure state outgoing edges) fail to failure state's failure
- if (failure state is initial state) use down arrow to denote
- if (failure state has no outgoing edges) fail to failure state's failure