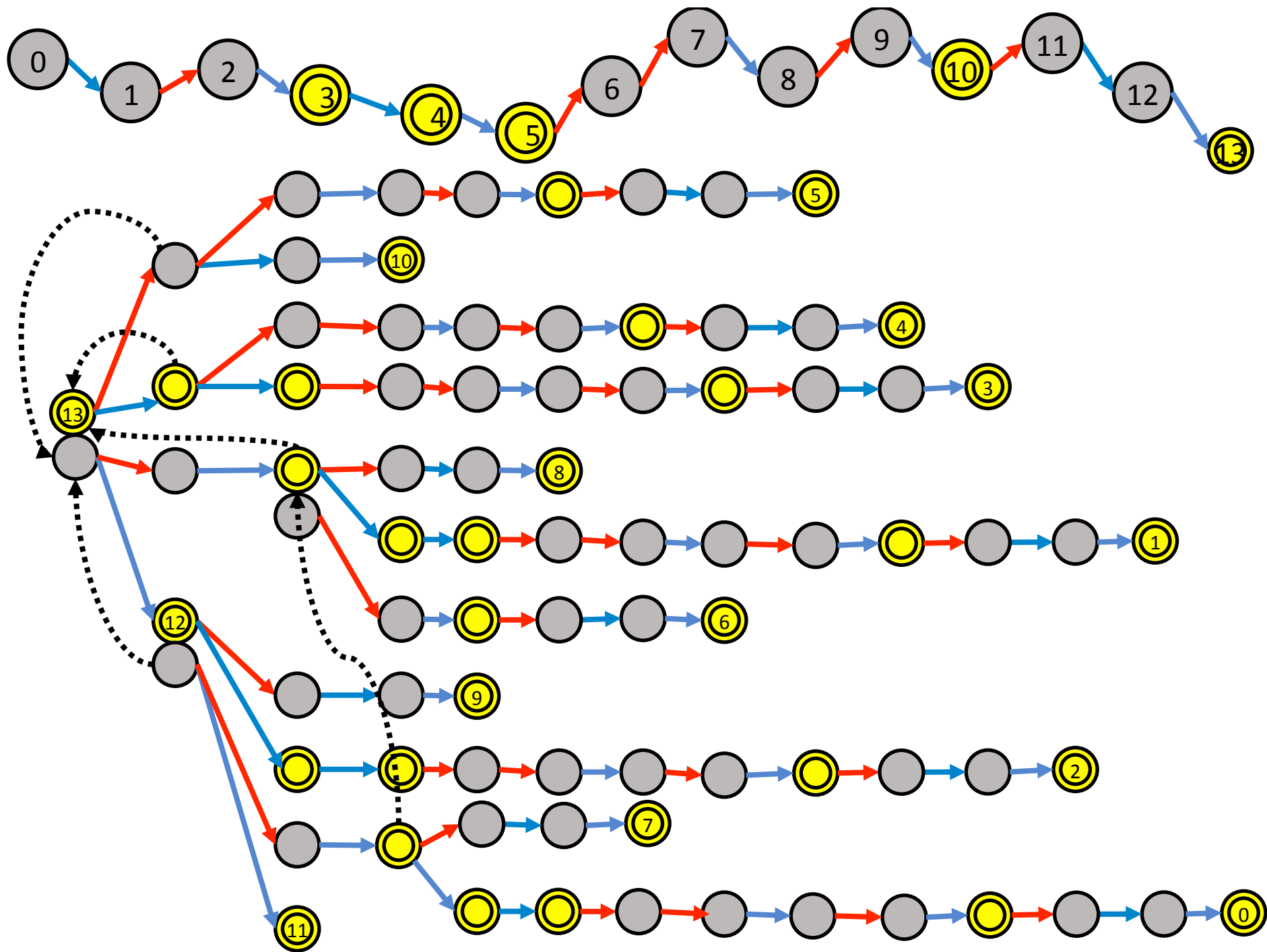
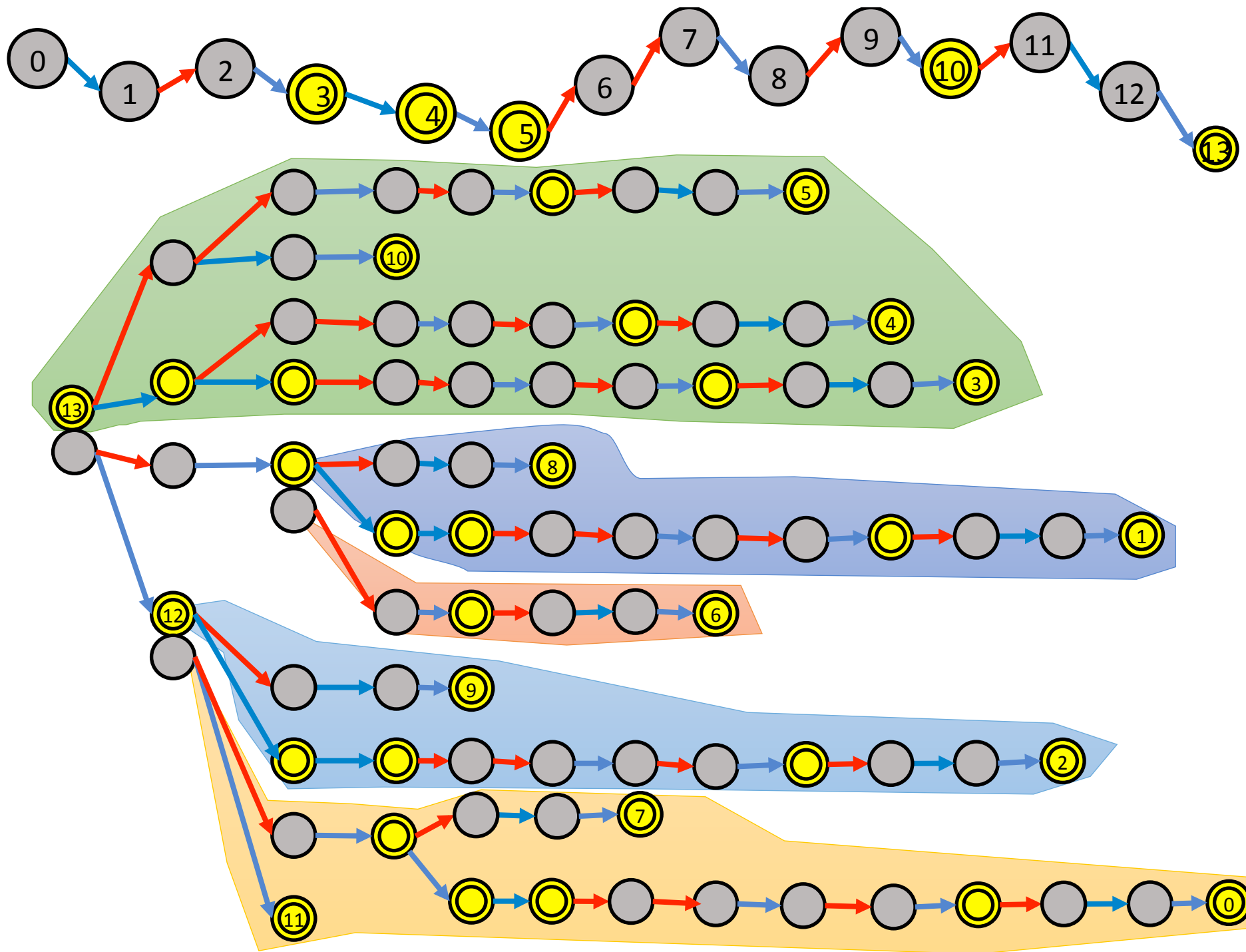


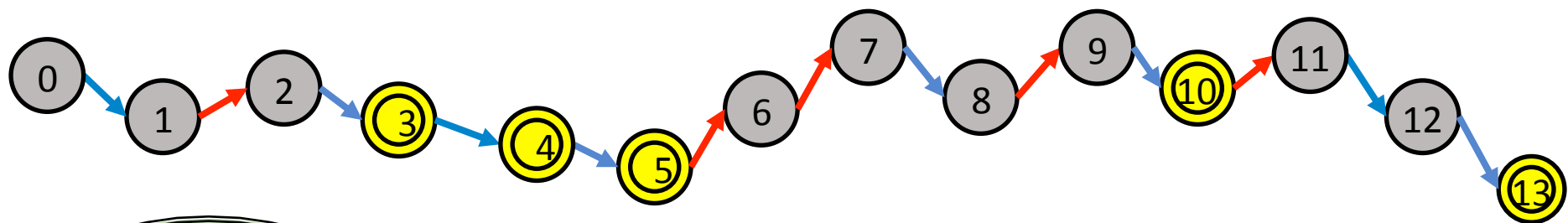
サフィックスツリーっぽいのを 作ってみる版 Mk. 2





部分列木の状態としての取り方

1. 大きい数字(suffix開始位置)から順に:
 - ① そのノード自身をふくめて、二重状態ノードが出るまで、根方向にのぼる。
 - ② (二重状態でも)根にたどりついたら直接の親ノードをえらび、これを根とする部分木を一つの状態と同一視する。状態のラベルは根のラベル、状態IDは部分木中のノードに振られた数字のうち最も若いもの。
 - ③ 根ではない二重状態にたどりついたら、ラベルを記憶し、そこからの部分木のノードに振られた数字を枚挙し、各数-1がついたノードからさかのぼった最初の二重状態ノード(または根)のラベルを比べ同じものを探す。見つかった場合、そのノードの部分木および共通祖先部分を同一視する部分木を同一状態とする。これを根にたどり着くまで繰り返す。
 - ④ 状態として同一視する部分木に含まれた数字は、繰り返して順にチェックするものから除外する。
2. 状態のリストができたなら、文字列の先頭から遷移をつけていく。すでに文字に対し遷移がある場合はそれに従い、ない場合は部分木のもつ数字のリストで遷移先を選び、最後の文字まで行い完成させる。



3,4,5,10,13

←13自身を根とする部分木から決定.

2, 9, 12,
1, 8

←まず12自身を根とする部分木からきまる.
11からさかのぼった最初の二重ラベルノードはラベル違い.
1と8からさかのぼった二重ラベルノードを根とする部分木,
二つの部分木の共通祖先は根になるので、決定.

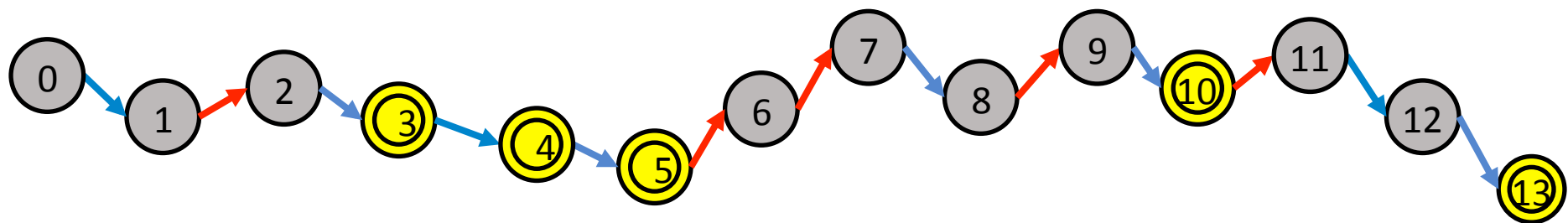
0,7,11,
6

←11からさかのぼった最初の二重ラベルノードと、
10はラベル違い、0はその前がないので、
6からさかのぼって最初の二重ラベルノードが一致、
共通祖先部分が根を含むので、決定.

もうつきたので、終了.

状態遷移の追加方法

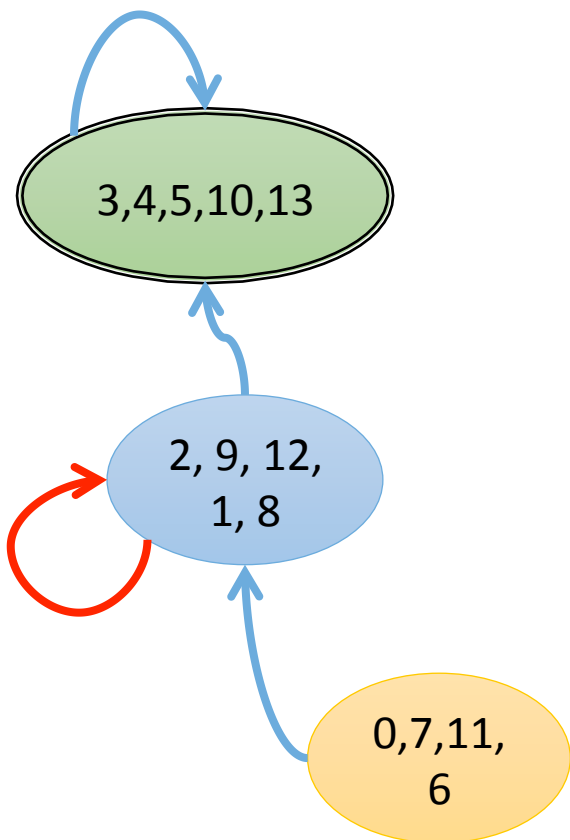
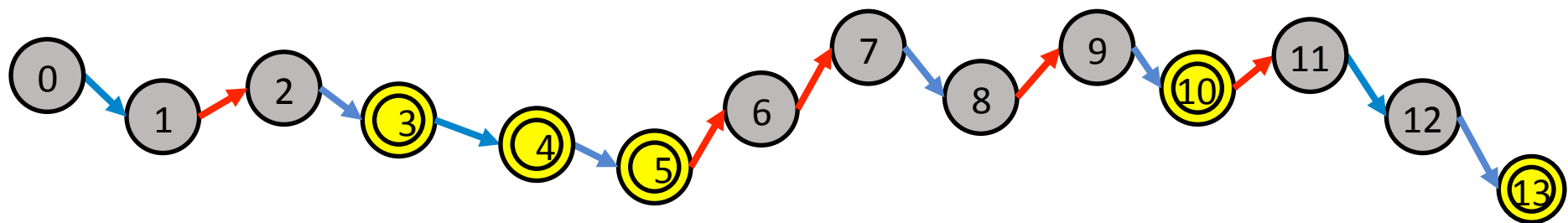
- ① 文字列の前からつけていく.
- ② 遷移がすでにある場合は、それに従う.
次の状態に該当する番号がなくてもかまわない.
- ③ ない場合、次の状態に該当する番号が含まれる
状態に遷移をつける.

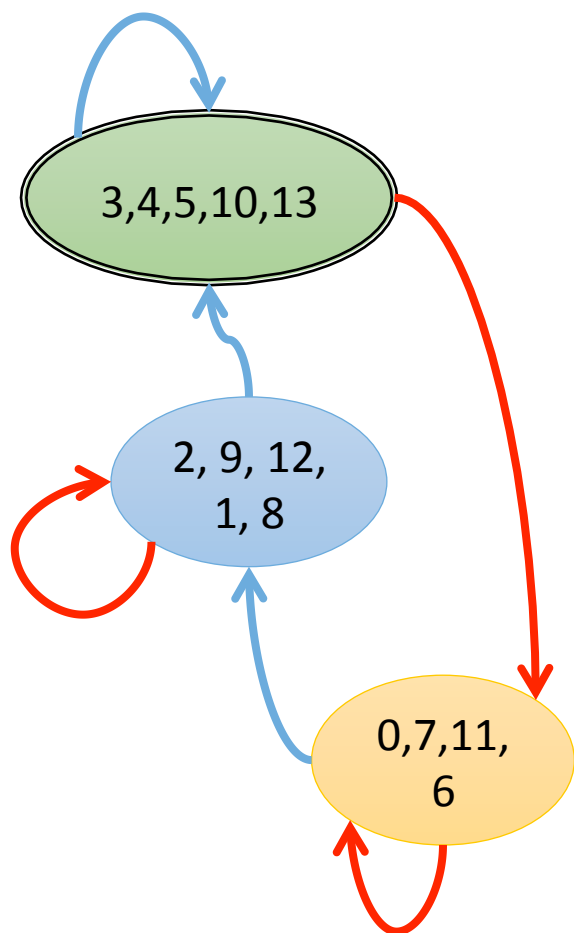
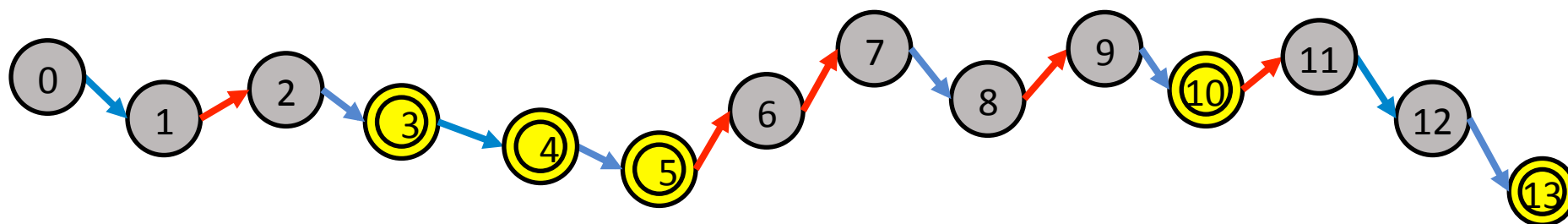


3,4,5,10,13

2, 9, 12,
1, 8

0,7,11,
6





おわっちゃった.

まとめ

- ① 部分列の反復、接尾辞の性質としての位置ズレして再出現する木の構造、などに気づかせてくれた。
- ② Suffix tree の線形時間構成法を参考にとすると、オーダーがおちるのでは、などの期待もふくらむ。
- ③ 接尾辞木ができていると全体をみわたせるが、全部でてきちゃってるので、オートマトン作成手続きを表現するのが面倒。
- ④ なので手続きとしてはトライ林でやるほうがシンプルだと自己満足。