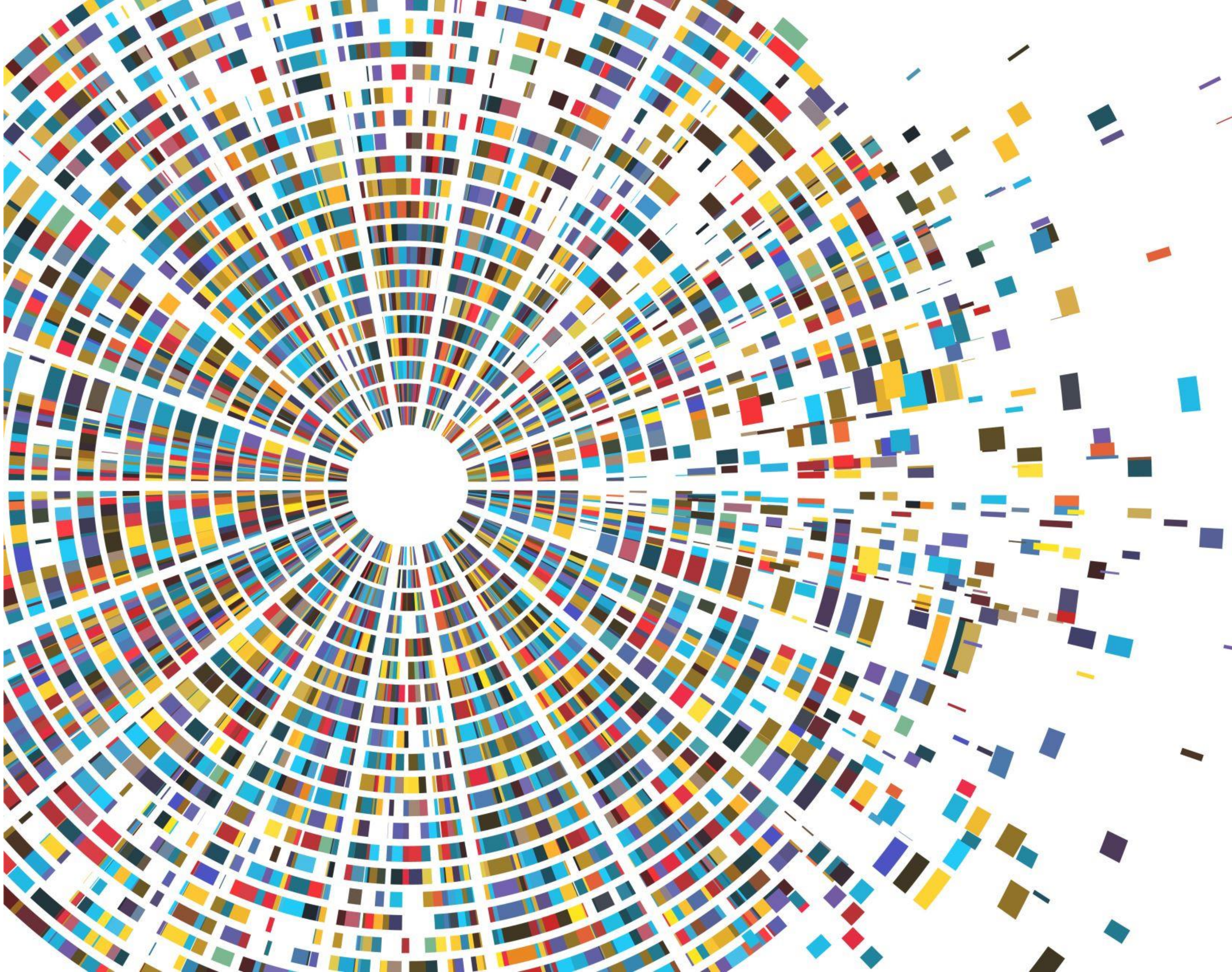
ТЕОРИЯ АВТОМАТОВ
Решенный тренировочный вариант контрольной работы
Задача 11

Направл.: Математика и компьютерные науки

к.ф.-м.н., доцент
Нагребецкая Ю.В.

Участвовал в проекте
студент гр. МЕН-290201
Зарипов В.

2022



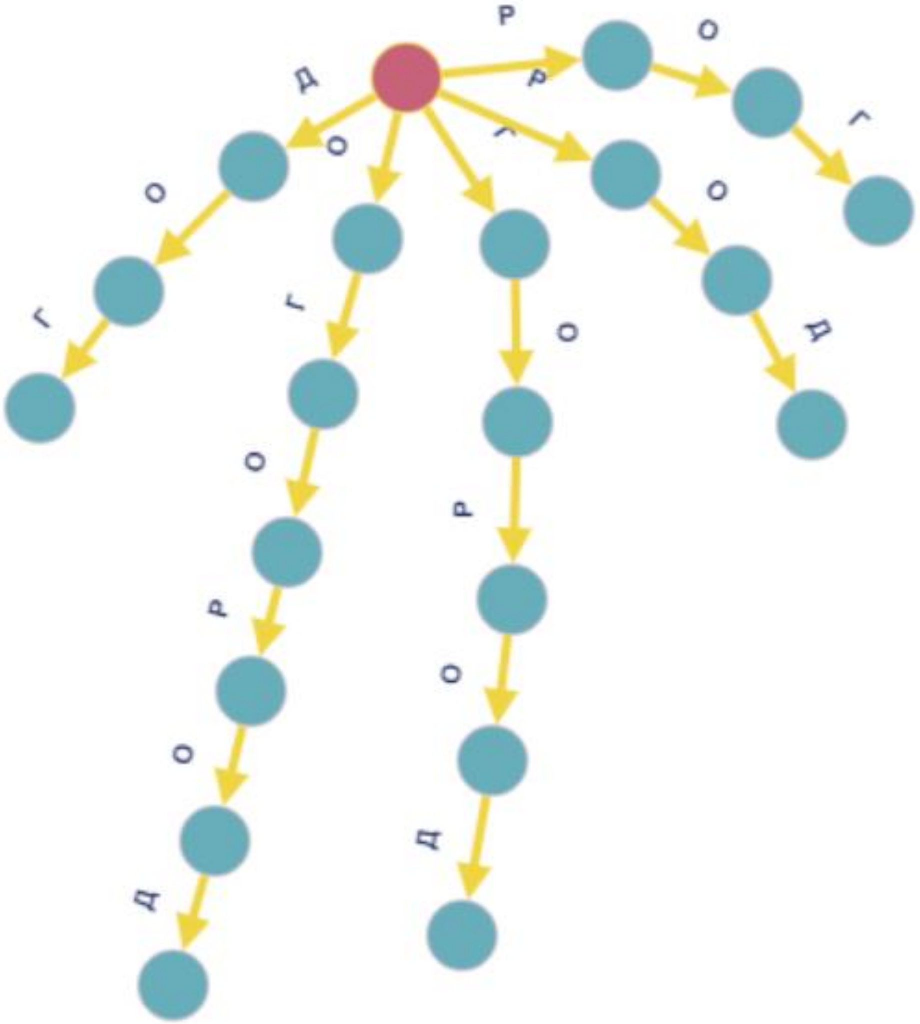
Задача 11

Построить автомат *Ахо-Корасик* для следующего набора шаблонов {ГОРОД, ОГОРОД, РОД, ДОГ, РОГ}

Построение бора (префиксного дерева, неполного автомата Ахо-Корасик)

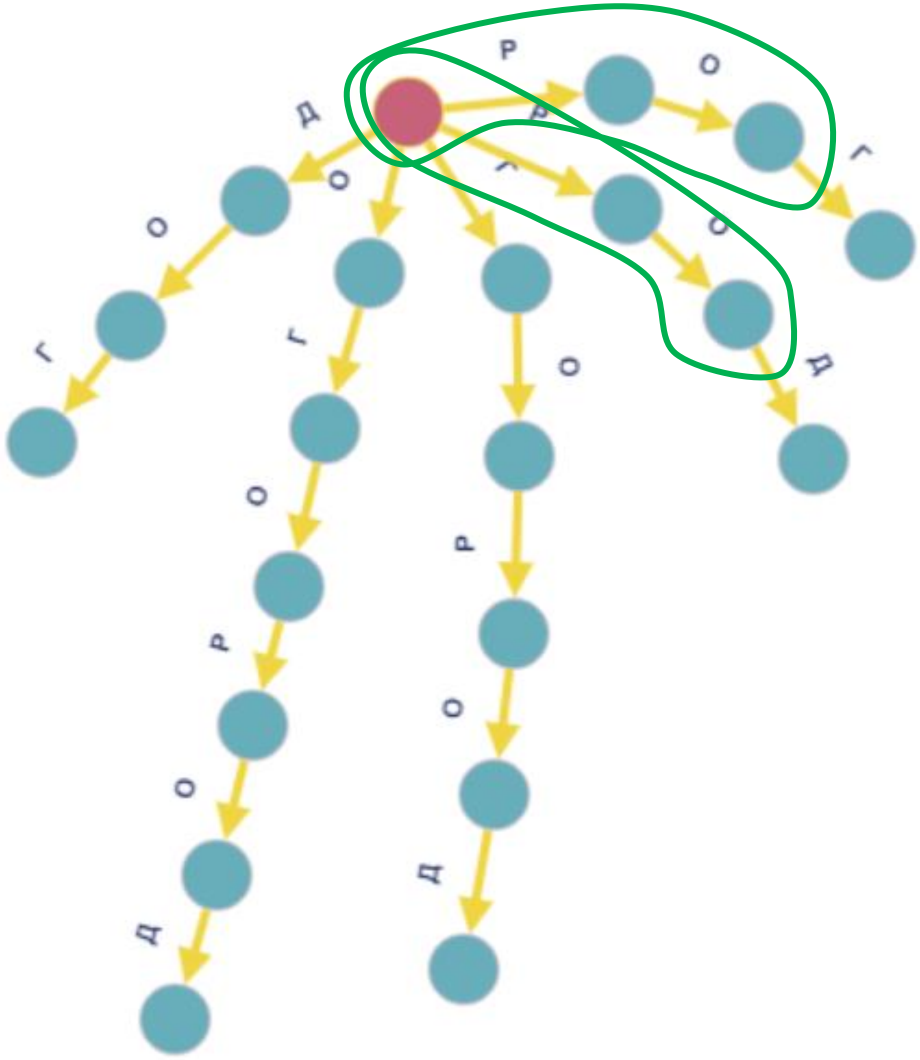
Алфавит: $\Sigma = \{Г, Р, О, Д\}$.

1) Построим **корневое ориентированное, помеченное** буквами алфавита, **дерево**, ветвями которого являются **шаблонные** слова.



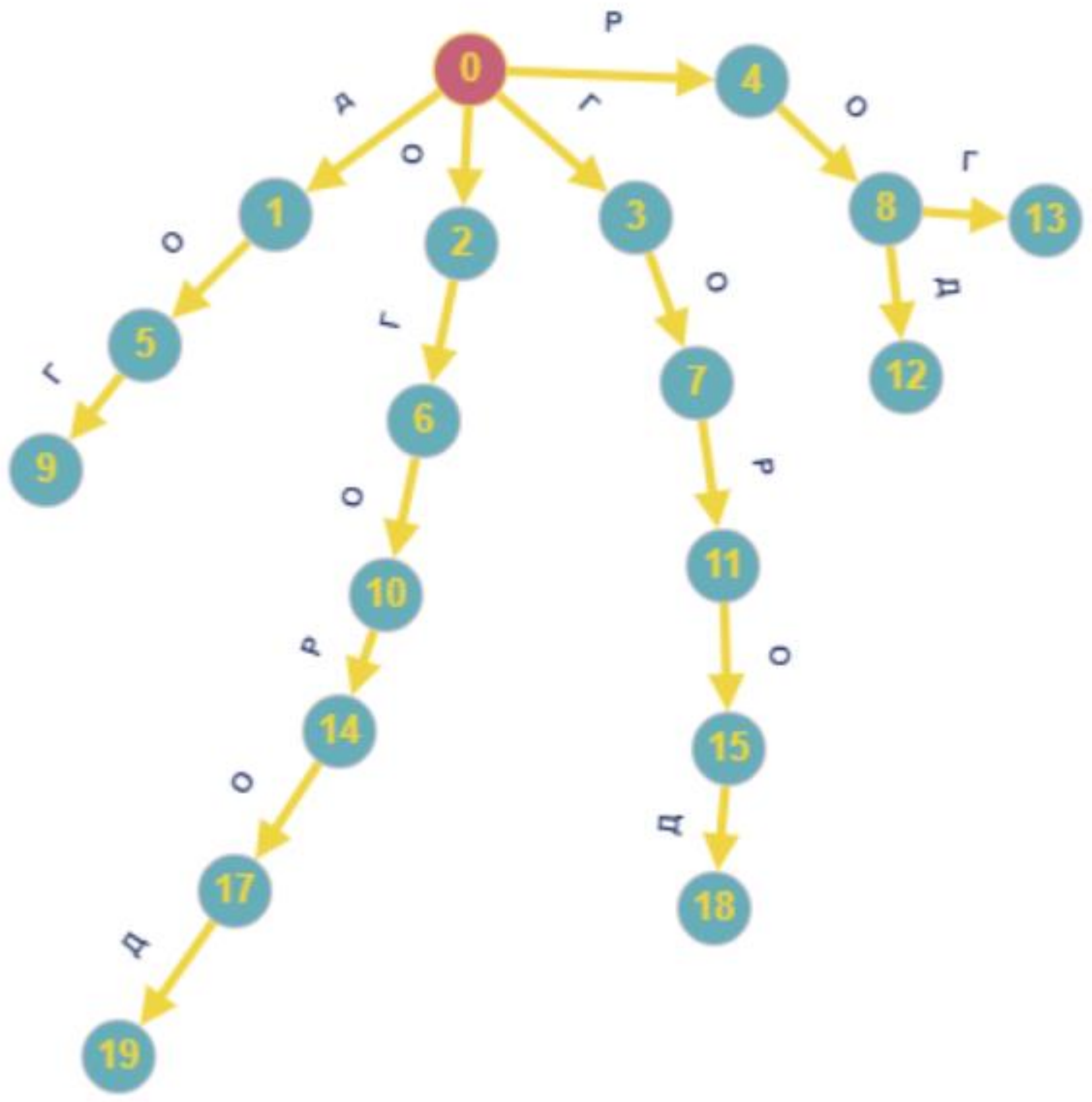
Построение бора (префиксного дерева)

2) Если у **шаблонных слов** есть общие **префиксы**, склеиваем их в одну цепь: РО-Д, РО-Г.



Построение бора (префиксного дерева)

Получаем дерево, которое называем **бор**.



Построение бора (префиксного дерева)

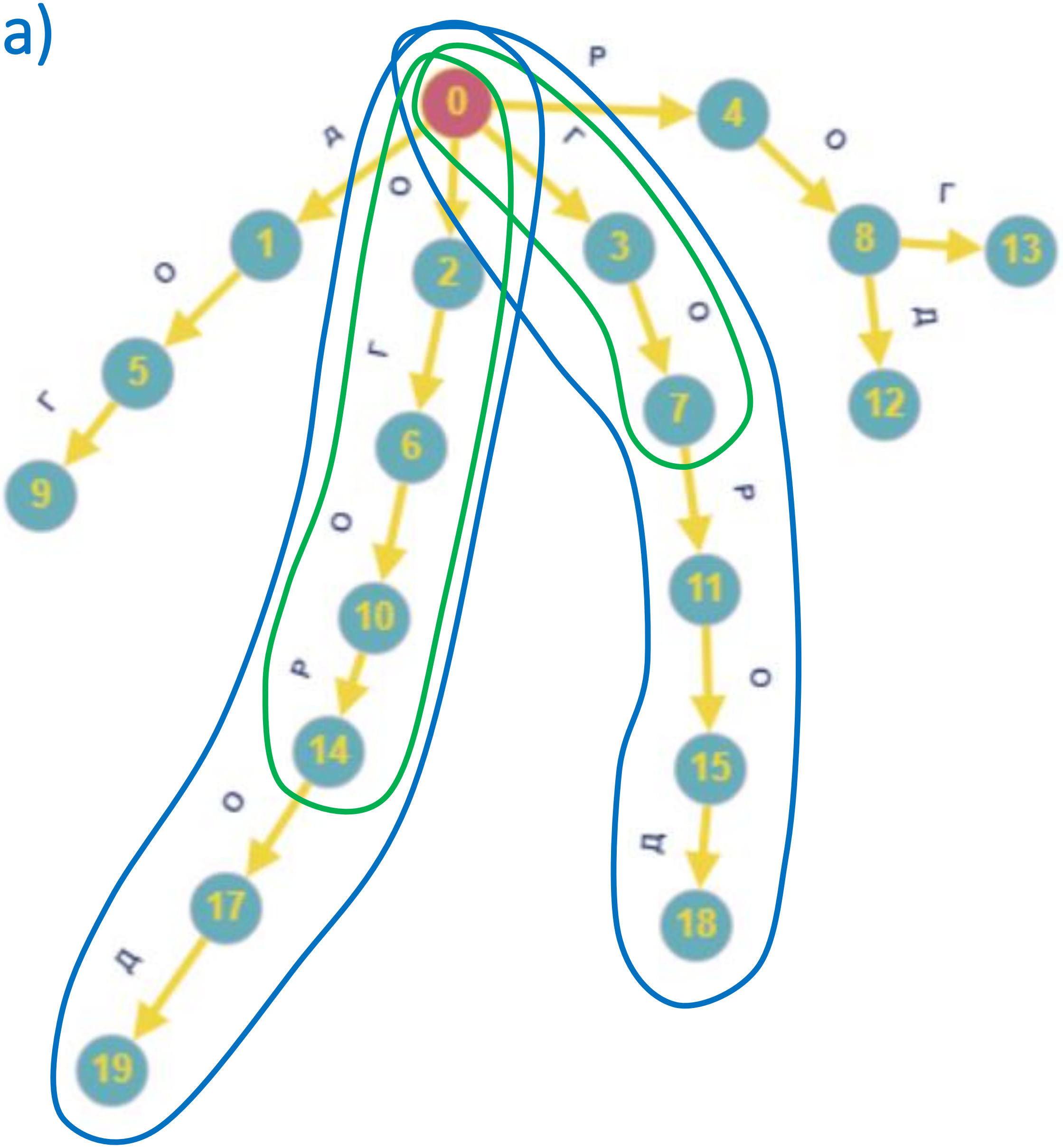
Пусть v – вершина дерева (бора), $f(v)$ – отец вершины v.

Будем строить **суффиксные ссылки** $\pi(v)$ для каждой вершины v.

Почему они называются **суффиксными**?

Каждую вершину v можно отождествить со словом v', читаемым от корня до этой вершины – **префиксом** некоторого **шаблонного** слова.

Например, для v=**7** слово v'=**ГО** - **префикс шаблонного** слова **ГОРОД**, **а для** v=**14** слово v'=**ОГОР** - **префикс шаблонного** слова **ОГОРОД.**

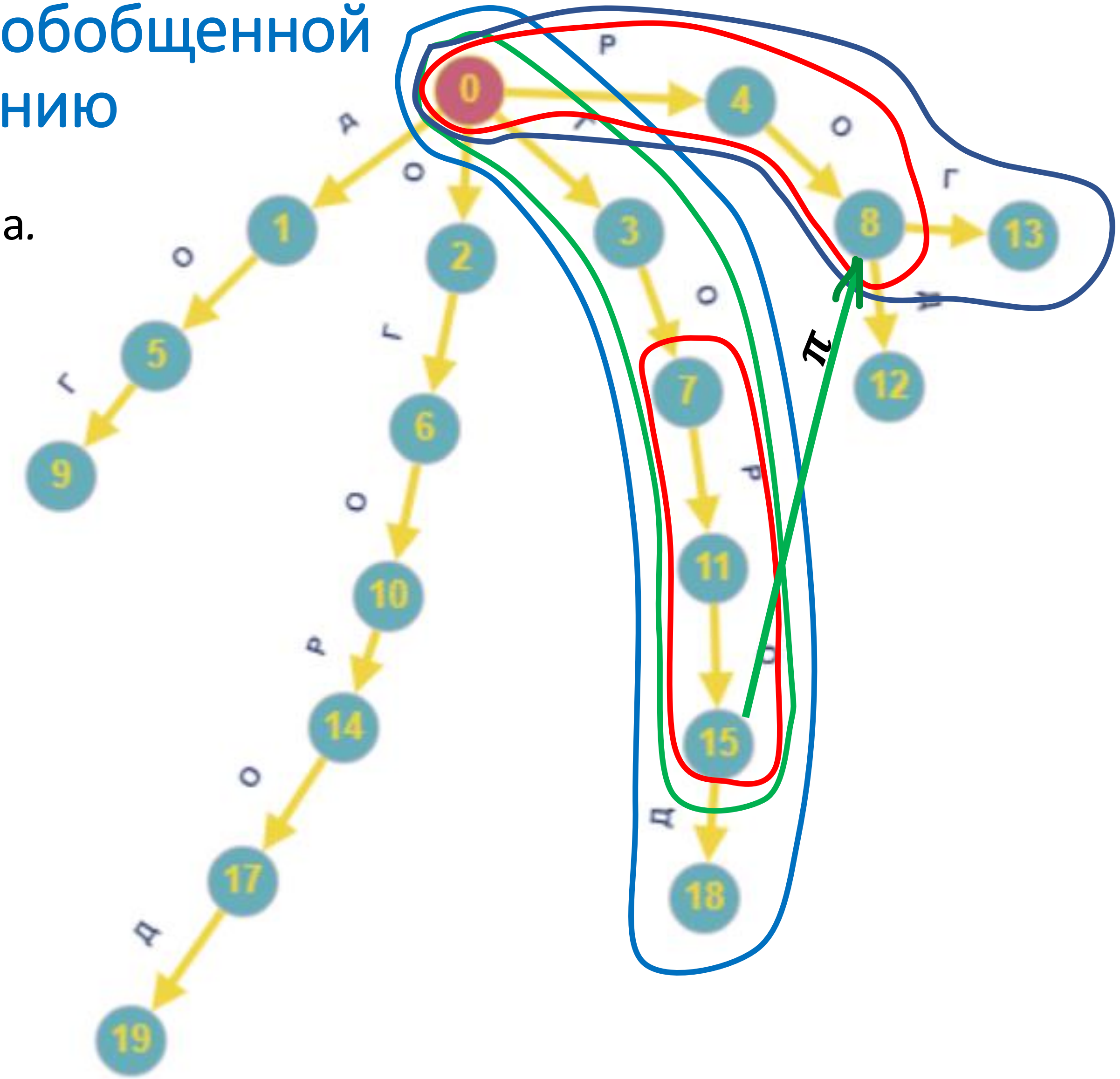


Формирование суффиксной ссылки (функции ошибки, функции отката, обобщенной префиксной функции) по определению

Пусть *v′*- **префикс** некоторого **шаблонного** слова.

Положим ***π*(*v*) =*w* (*π*(*v′*) =*w′*),**
если *w′* - **максимальный суффикс префикса** *v′*, *w′*≠ *v′*, такой,
что *w′* - **префикс** некоторого
(не обязательно другого)
шаблонного слова.

Например, ***π*(15) =8 (*π*(ГОРО) = РО),**
поскольку **префикс** *v′*=**ГОРО** шаблонного
слова **ГОРОД** имеет **суффикс (максимальный)** *w′* =**РО,** который
в свою очередь является префиксом
другого шаблонного слова **РОГ** **(а также шаблонного слова РОД).**



Алгоритм формирования суффиксной ссылки (функции ошибки, функции отката, обобщенной префиксной функции)

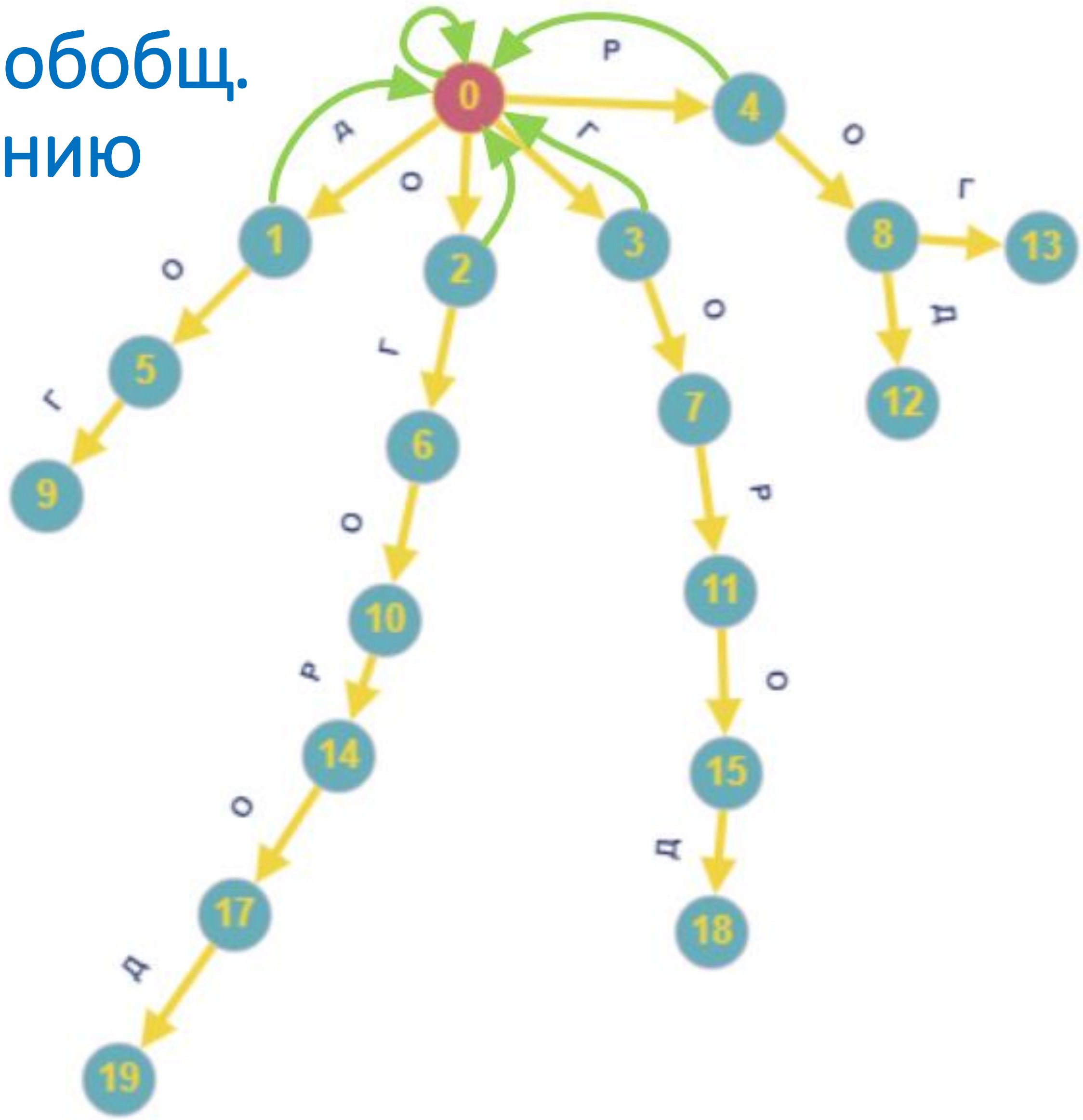
Идем по суффиксной ссылке родителя.

Если из этого состояния можно перейти по текущему символу, то устанавливаем суффикскную ссылку на
полученное состояние.

Если из состояния по суффиксной ссылке нельзя сделать переход, то продолжаем подниматься по бору по суффиксным ссылкам родителей.

Чтобы гарантированно у родителя была суффиксная ссылка используется алгоритм поиска в ширину от корневой вершины.

Формирование суффиксной ссылки (функции ошибки, функции отката, обобщ. префиксной функции) по определению



3) Начинаем обходить бор в поиском в ширину.
Полагаем **$\pi(0)$ =0 или, что тоже самое, $\pi(\lambda) = \lambda$**

4) Вершины 1,2,3,4 отстоят от корня 0 на расстоянии 1.
Полагаем **$\pi(1)$ =0, $\pi(2)$ =0, $\pi(3)$ =0, $\pi(4)$ =0**
или, что тоже самое,
π(Г) =π(О) =π(Р) =π(Г) = λ

Действительно, например, для префикса *v'*=**Г** **шаблонного слова ГОРОД**, только суффикс $w' = \lambda$ этого префикса может быть префиксом других шаблонных слов: **ОГОРОД, РОД, РОГ.**

Формирование суффиксной ссылки (по определению)

5) Вершины 5,6,7,8 отстоят от корня 0 на расстоянии 2.
Полагаем **$\pi(5) = 2$, $\pi(6) = 3$, $\pi(7) = \pi(8) = 2$,**
или, что тоже самое,
π(ДО) =О, π(ОГ) =Г, π(ГО) = π(РО) =О.

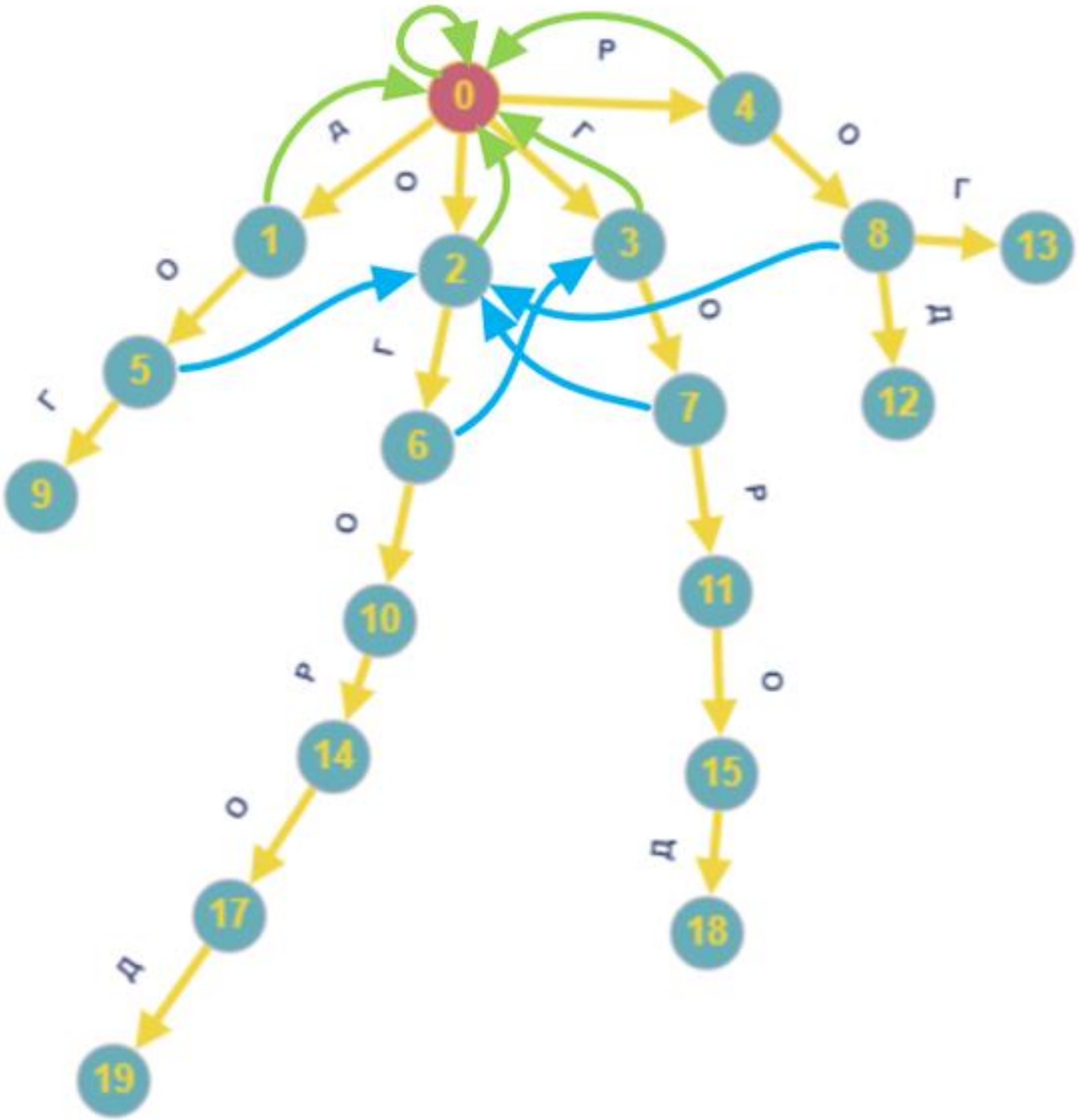
Действительно,

5.1) для **префикса** *v'*=**ДО шаблонного слова ДОГ максим**. **суффикс** w'= **О** этого **префикса** является **префиксом шаблонного** слова: **ОГОРОД;**

5.2) для **префикса** *v'*=**ОГ шаблонного слова ОГОРОД максим**. **суффикс** w'= **Г** этого **префикса** является **префиксом шаблонного** слова: **ГОРОД;**

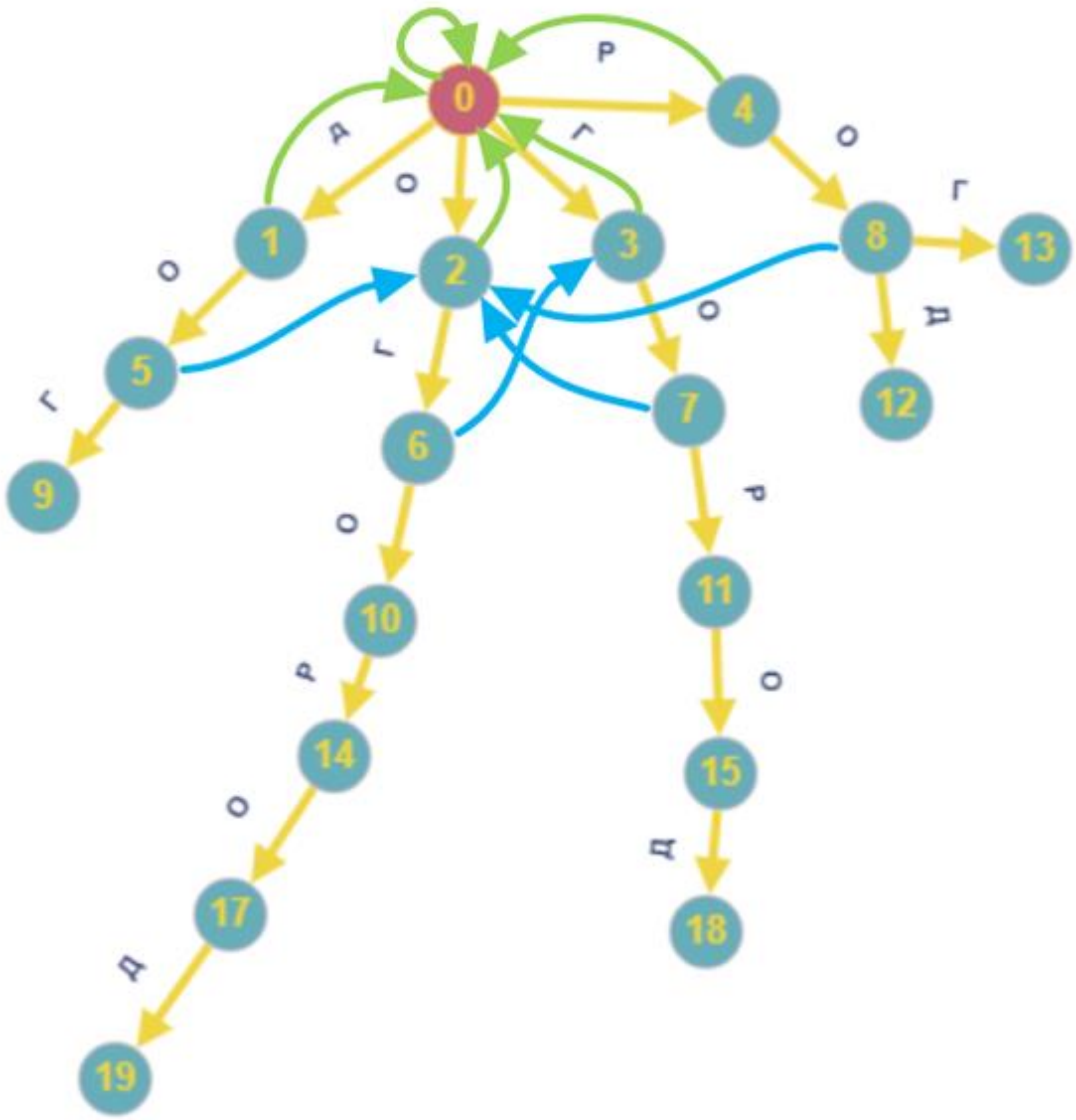
5.3) для **префикса** *v'*=**ГО шаблонного слова ГОРОД максим**. **суффикс** w'= **О** этого **префикса** является **префиксом шаблонного слова**: **ОГОРОД;**

5.4) для **префикса** *v'*=**РО шаблонного слова РОД максим**. **суффикс** w'= **О** этого **префикса** является **префиксом шаблонного слова**: **ОГОРОД;**



Формирование суффиксной ссылки по алгоритму

5) Вершины 5,6,7,8 отстоят от корня 0 на расстоянии 2.
Полагаем ***π*(5) =2, *π*(6) =3, *π*(7) = *π*(8) =2,**
или, что тоже самое,
***π*(ДО) =О, *π*(ОГ) =Г, *π*(ГО) = *π*(РО) =О.**



Действительно,

5.1) в вершину **5** мы попадаем из **родительской** вершины **1** по букве **О**, ***π*(1)=0,** и из вершины **0** по букве **О** есть переход в вершину **2**. Поэтому имеем ***π*(5) =2.**

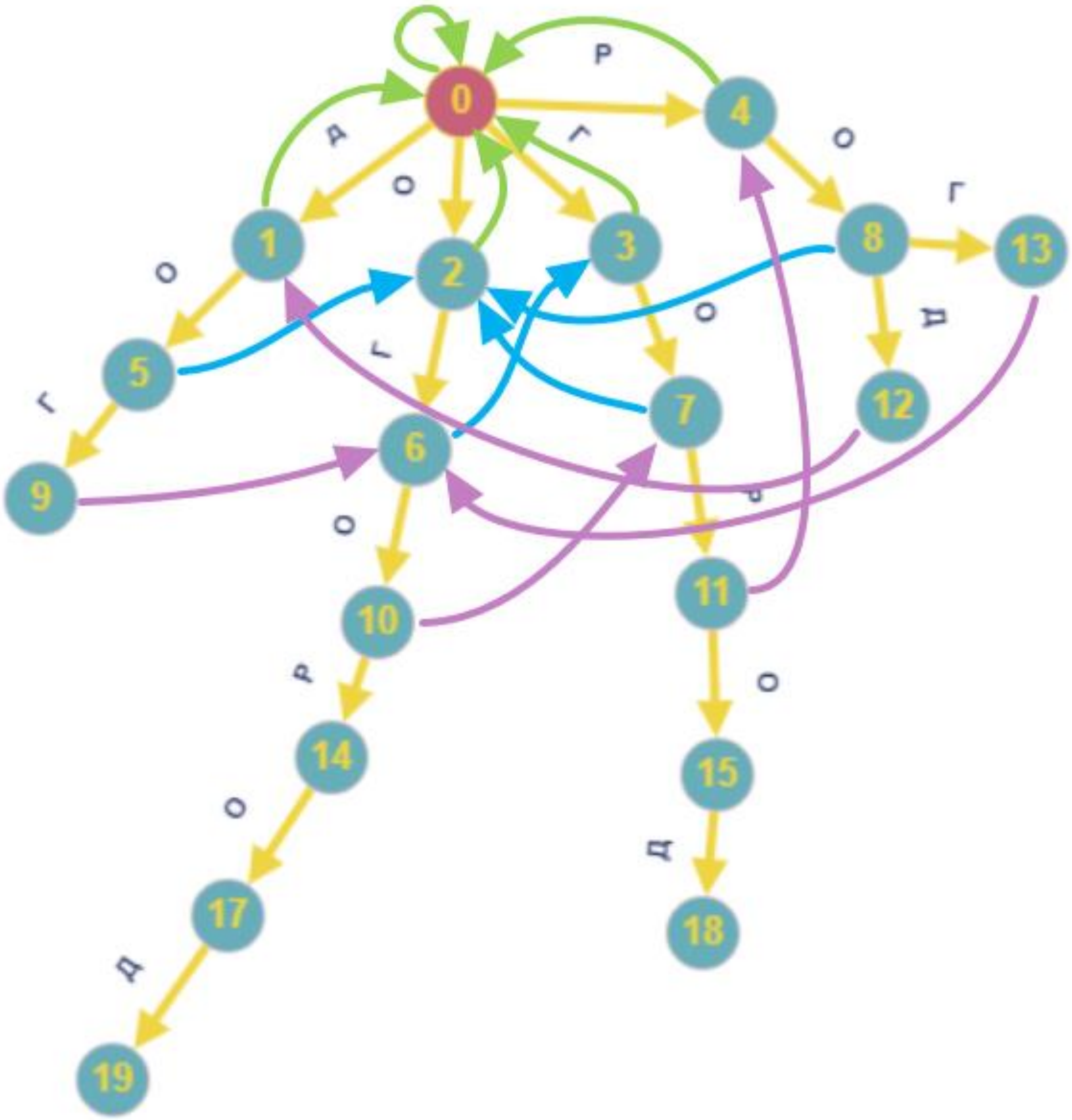
5.2) в вершину **6** мы попадаем из **родительской** вершины **2** по букве **Г**, ***π*(2)=0,** и из вершины **0** по букве **Г** есть переход в вершину 3. Поэтому полагаем ***π*(6) =3.**

5.3) $7 \xleftarrow{\text{О}} 3,\ \pi(3) = 0,\ 0 \xrightarrow{\text{О}} 2 \Rightarrow \pi(7) = 2$

5.4) $8 \xleftarrow{\text{О}} 4,\ \pi(4) = 0,\ 0 \xrightarrow{\text{О}} 2 \Rightarrow \pi(8) = 2$

Формирование суффиксной ссылки по определению

6) Вершины 9,10,11,12,13 отстоят от корня 0 на расстоянии 3.
Полагаем **π(9) =6, π(10) =7, π(11) = 4, π(12) =1, π(13) =6**
или, что тоже самое,
π(ДОГ) =ОГ, π(ОГО) =ГО, π(ГОР) =Р, π(РОД) =Д, π(РОГ) =ОГ.



Действительно,
6.1) для **префикса** *v'*=**ДОГ шаблонного слова ДОГ**
максим. **суффикс** w'= **ОГ** этого **префикса** является **префиксом шаблонного слова**: **ОГОРОД;**
6.2) для **префикса** *v'*=**ОГО шаблонного слова ОГОРОД**
максим. **суффикс** w'= **ГО** этого **префикса** является **префиксом** шаблонного слова: **ГОРОД;**
6.3) для **префикса** *v'*=**ГОР шаблонного слова ГОРОД**
максим. **суффикс** w'= **Р** этого **префикса** является **префиксом** шаблонного слова: **РОД;**
6.4) для **префикса** *v'*=**РОД шаблонного слова РОД**
максим. **суффикс** w'= **Д** этого **префикса** является **префиксом шаблонного слова**: **ДОГ;**
6.5) для **префикса** *v'*=**РОГ шаблонного слова РОГ**
максим. **суффикс** w'= **ОГ** этого **префикса** является **префиксом шаблонного слова**: **ОГОРОД.**

Формирование суффиксной ссылки по алгоритму

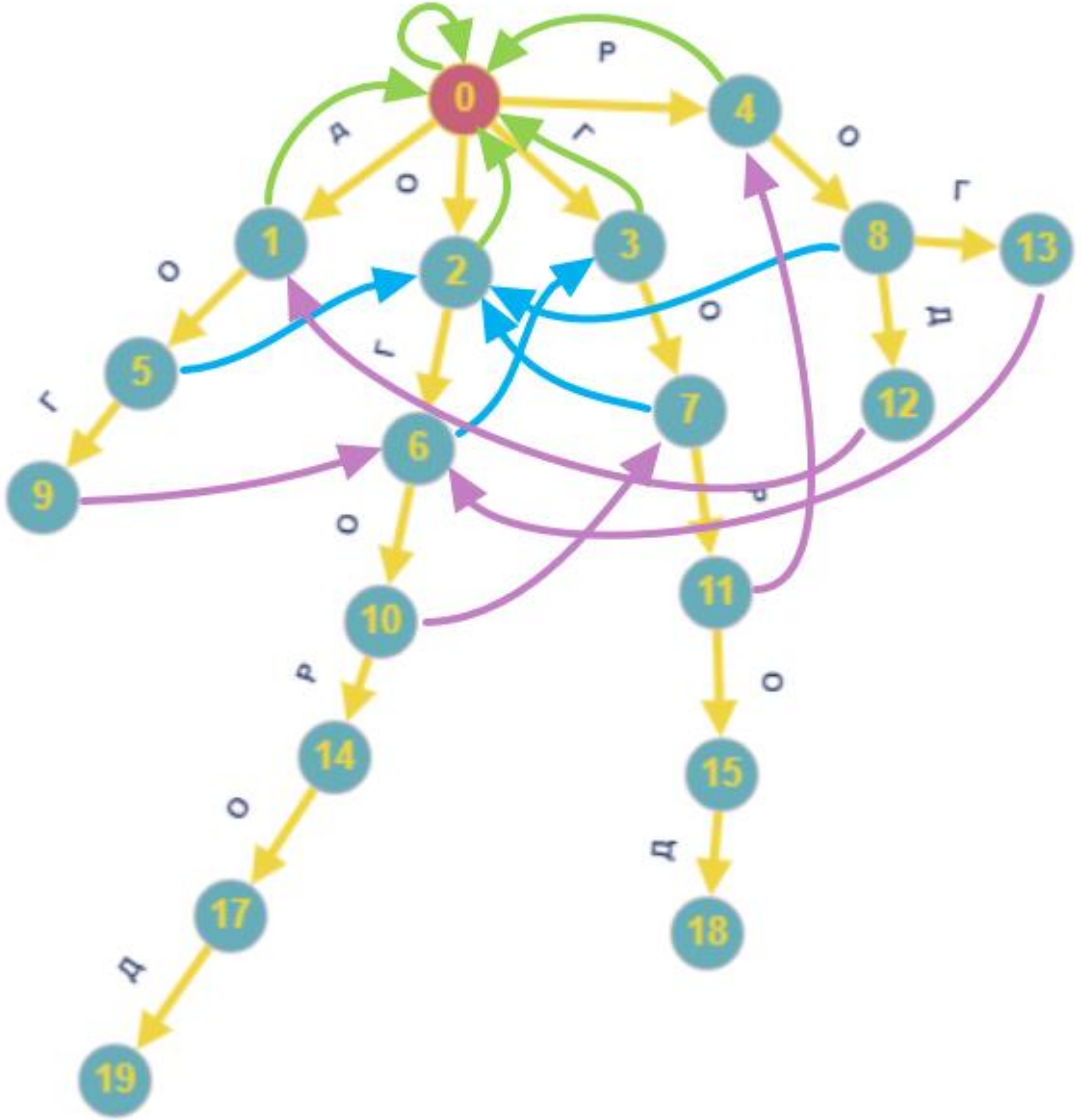
6) Вершины 9,10,11,12,13 отстоят от корня 0 на расстоянии 3.
Полагаем ***π*(9) =6, *π*(10) =7, *π*(11) = 4, *π*(12) =1, *π*(13) =6**
или, что тоже самое,
***π*(ДОГ) =ОГ, *π*(ОГО) =ГО, *π*(ГОР) =Р, *π*(РОД) =Д, *π*(РОГ) =ОГ.**

$$6.1)\quad 9 \xleftarrow{\text{Г}} 5,\ \pi(5) = 2,\ 2 \xrightarrow{\text{Г}} 6 \Rightarrow \pi(9) = 6$$

$$6.2)\quad 10 \xleftarrow{\text{О}} 6,\ \pi(6) = 3,\ 3 \xrightarrow{\text{О}} 7 \Rightarrow \pi(10) = 7$$

6.3) В вершину **11** мы попадаем из ее **родительской** вершины **7** по букве **Р**, ***π*(7)=2,** но из вершины **2** по букве **Р** НЕТ перехода. Поэтому поднимаемся выше по ссылкам: ***π*(2)=0** и из вершины 0 ЕСТЬ переход по букве **Р** в вершину **4**, следовательно, ***π*(11)=4.**

6.4) Аналогично, в вершину **12** мы попадаем из ее **родительской** вершины **8** по букве **Д**, ***π*(8)=2,** но из вершины **2** по букве **Д** НЕТ перехода. Поэтому поднимаемся выше по ссылкам: ***π*(2)=0** и из вершины 0 ЕСТЬ переход по букве **Д** в вершину **1**, следовательно, ***π*(12)=1.**



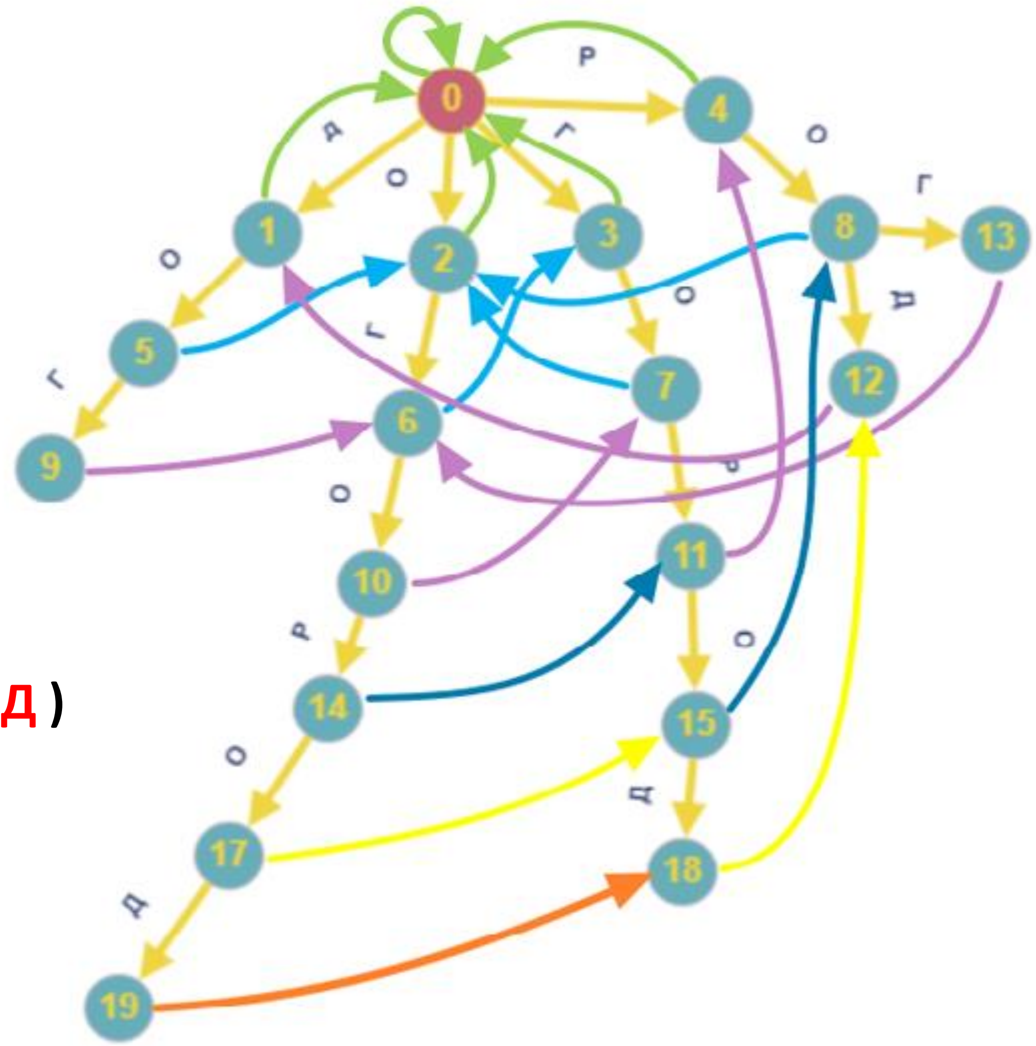
Формирование суффиксной ссылки по алгоритму

7) Аналогичным образом формируем все оставшиеся суффиксные ссылки

π(14) =11, π(15) =8 (π(ОГОР) =ГОР, π(ГОРО) =РО)

π(17) = 15, π(18) =12 (π(ОГОРО) =ГОРО, π(ГОРОД) =РОД)

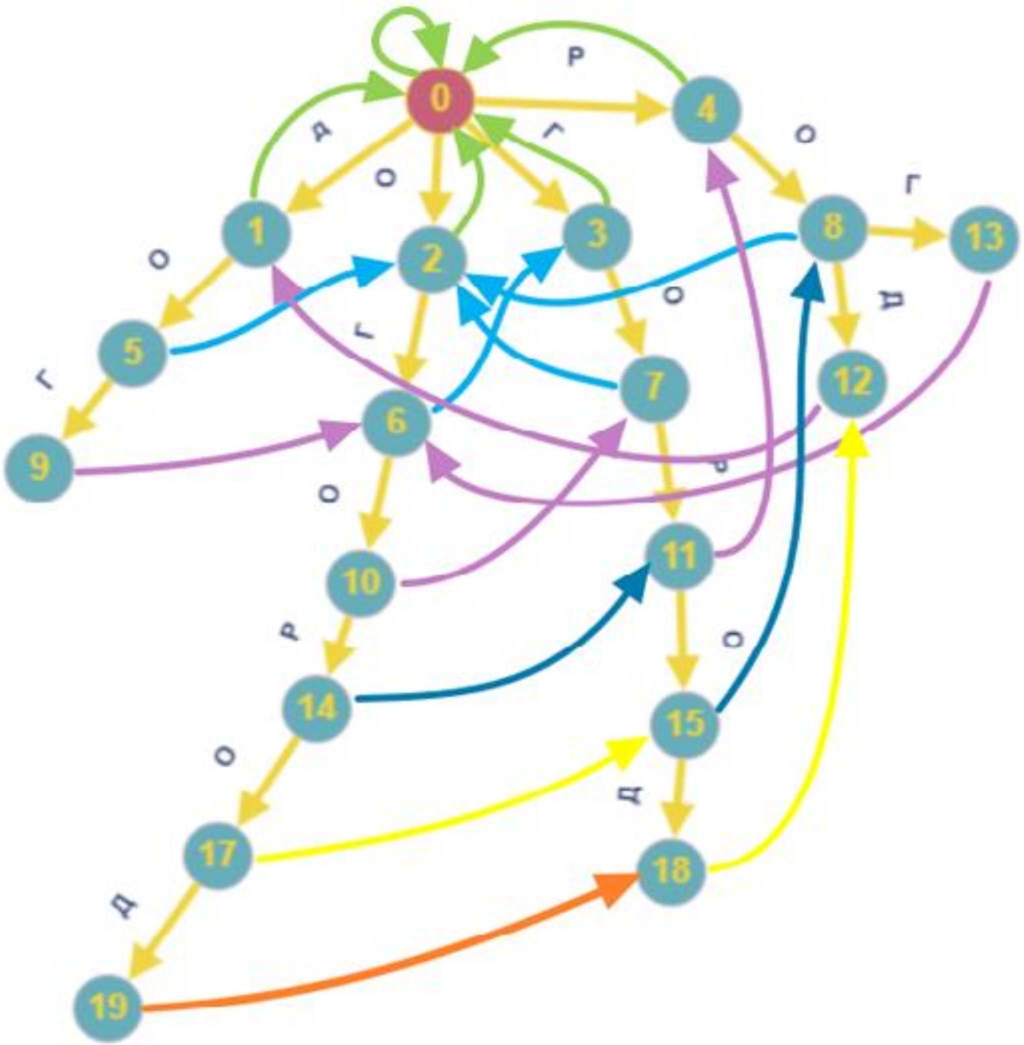
π(19) =18 (π(ОГОРОД) =ГОРОД)



Формирование суффиксной ссылки по алгоритму

Приведем таблицу суффиксных ссылок

v	$\pi(v)$
0	0
1	0
2	0
3	0
4	0
5	2
6	3
7	2
8	2
9	6
10	7
11	4
12	1
13	6
14	11
15	8
17	15
18	12
19	18

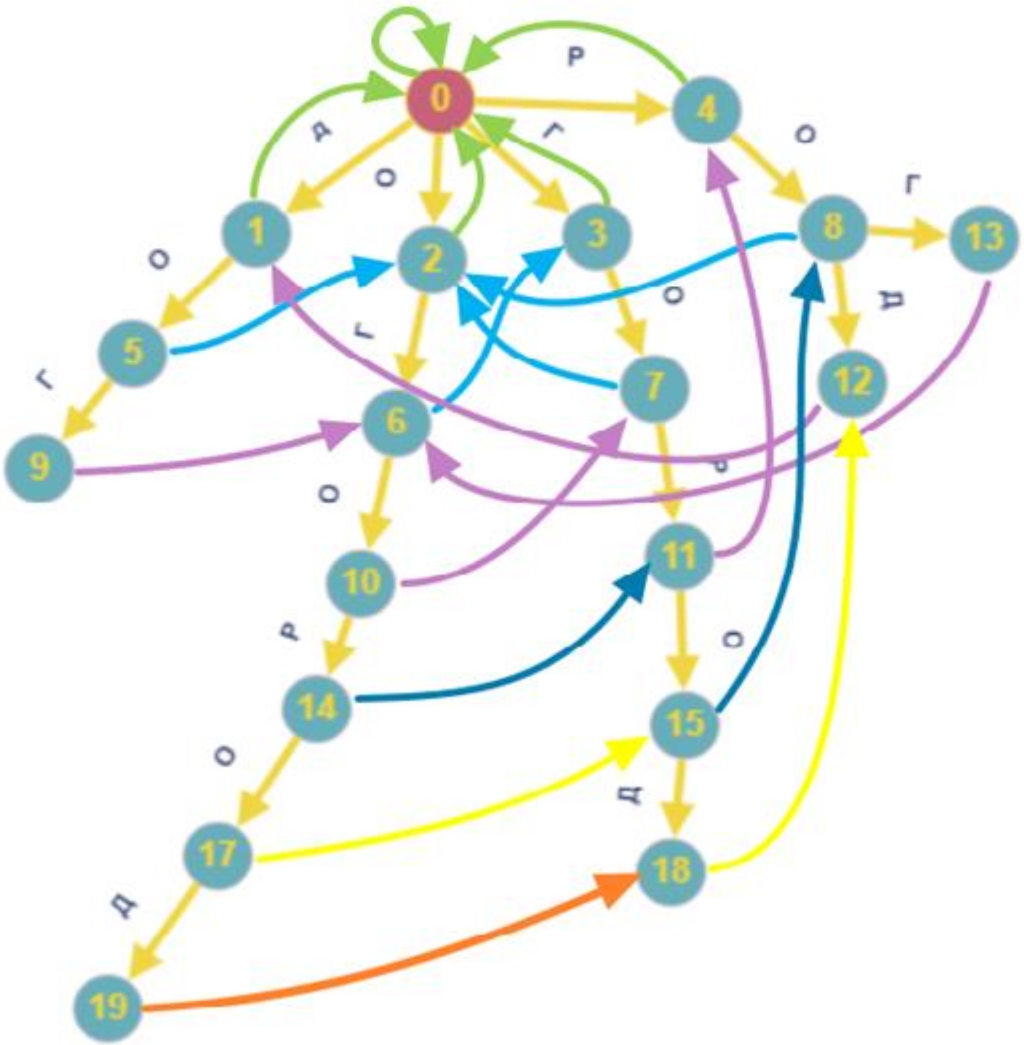


Формирование суффиксной ссылки по алгоритму

Приведем таблицу суффиксных ссылок

v	π(v)
λ	λ
Д	λ
О	λ
Г	λ
Р	λ
ДО	О
ОГ	Г
ГО	О
РО	О
ДОГ	ОГ
ОГО	ГО
ГОР	Р

v	π(v)
РОД	Д
РОГ	ОГ
ОГОР	ГОР
ГОРО	РО
ОГОРО	ГОРО
ГОРОД	РОД
ОГОРОД	ГОРОД



Формирование суффиксной ссылки по алгоритму (протокол)

Приведем протокол формирования суффиксных ссылок (функции ошибки, функции отката)

v	π(v) (I шаг)	π(v) (II шаг)	π(v) (III шаг)	π(v) (IV шаг)	π(v) (V шаг)	π(v) (VI шаг)
λ	λ	λ	λ	λ	λ	λ
Д	λ	λ	λ	λ	λ	λ
О	λ	λ	λ	λ	λ	λ
Г	λ	λ	λ	λ	λ	λ
Р	λ	λ	λ	λ	λ	λ
ДО		О	О	О	О	О
ОГ		Г	Г	Г	Г	Г
ГО		О	О	О	О	О
РО		О	О	О	О	О
ДОГ			ОГ	ОГ	ОГ	ОГ
ОГО			ГО	ГО	ГО	ГО
ГОР			Р	Р	Р	Р

Формирование суффиксной ссылки по алгоритму (протокол)

Продолжение

v	π(v) (I шаг)	π(v) (II шаг)	π(v) (III шаг)	π(v) (IV шаг)	π(v) (V шаг)	π(v) (VI шаг)
РОД			Д	Д	Д	Д
РОГ			ОГ	ОГ	ОГ	ОГ
ОГОР				ГОР	ГОР	ГОР
ГОРО				РО	РО	РО
ОГОРО					ГОРО	ГОРО
ГОРОД					РОД	РОД
ОГОРОД						ГОРОД

Построение (полного) автомата Ахо-Корасик

Алфавит: $\Sigma = \{\text{Г}, \text{Р}, \text{О}, \text{Д}\}$ (он, впрочем, может быть расширен)

Начальная вершина – корень бора – вершина 0

Заключительные вершины – листы бора: вершины 9, 19, 18, 12, 13

Принимаемый (распознаваемый язык) – множество слов, суффиксами которых являются шаблонные слова: $\{\text{ДОГ}, \text{ОГОРОД}, \text{ГОРОД}, \text{РОД}, \text{РОГ}\}$
Начальная вершина – корень бора – вершина 0

Вершины автомата – вершины бора (1-19) (префиксы шаблонных слов)

Достроим при помощи суффиксн. ссылок **бор** (**неполный автомат Ахо-Корасик**) до (**полного**) **автомата Ахо-Корасик**

Алгоритм формирования таблицы переходов (полного) автомата Ахо-Корасик
Обходим бор от корня поиском в ширину.
Для любой вершины *v* бора и любой буквы *c* алфавита полагаем
a) *v.c:=u,* если в боре был переход *v.c=u;*
b) *Иначе* (если в боре не было перехода *v.c=u) полагаем v:=* π(*v*) *и перейти к пункту a),*
c) *для v.c:=* λ, если дошли до корня *v=* λ, а из него в боре нет перехода по букве *c*.
Алгоритм работает корректно, так как в худшем случае мы поднимемся до корня, а из корня есть переходы по букве алфавита или в корневую вершину.

Формирование таблицы переходов (полного) автомата Ахо-Корасик

8) Полагаем для автомата **0.Д=1, 0.О=2, 0.Г=3**, **0.Р=4**, так как в боре есть соответствующие переходы.
9) Далее,
9.1) **1.О=5**, поскольку в боре есть переход из вершины **1** по букве **О** в вершину **5 (δ(1,О)=5)**.
9.2) По букве **Д** из вершины **1** нет перехода.
Поднимаемся вверх по ссылкам, рассматриваем ***π*(1) =0,** из вершины **0** по букве **Д** тоже есть переход в вершину **1**, поэтому полагаем **1.Д= δ(*π*(1),О)= δ(0,О)= 1.**
9.3) По букве **Г** и **Р**, аналогично, нет переходов из вершины **1**, но есть переход из вершины **0** в вершины **3,4.**
Полагаем **1.Г= δ(*π*(1),Г)= δ(0,Г)=3, 1.Р = δ(*π*(1),Г)= δ(0,Р)=4.**
9.4) Аналогично, **2.Г=6**, **2.О=2, 2.Д**=**1**, **2.Р=4,**
3.О=7, **3.Г=3, 3.Д=1**, **3.Р=4, 4.О=8**, **4.Д=1, 4.Г=3**, **4.Р=4.**
10) Рассмотрим вершину **5**. Ясно, что **5.Г=9.**
По другим буквам из вершины **5** нет переходов, ***π*(5) =2,** но и из вершины 2 есть переход только по букве **Г.** Поэтому поднимаемся выше по ссылкам: ***π*(2) =0** и полагаем
5.Д= δ(*π*(*π*(5)),Д)= δ(0,Д)= 1, 5.О= δ(*π*(*π*(5)),О)= δ(0,О)= 2,
5.Р= δ(*π*(*π*(5)),Р)= δ(0,Р)= 4.

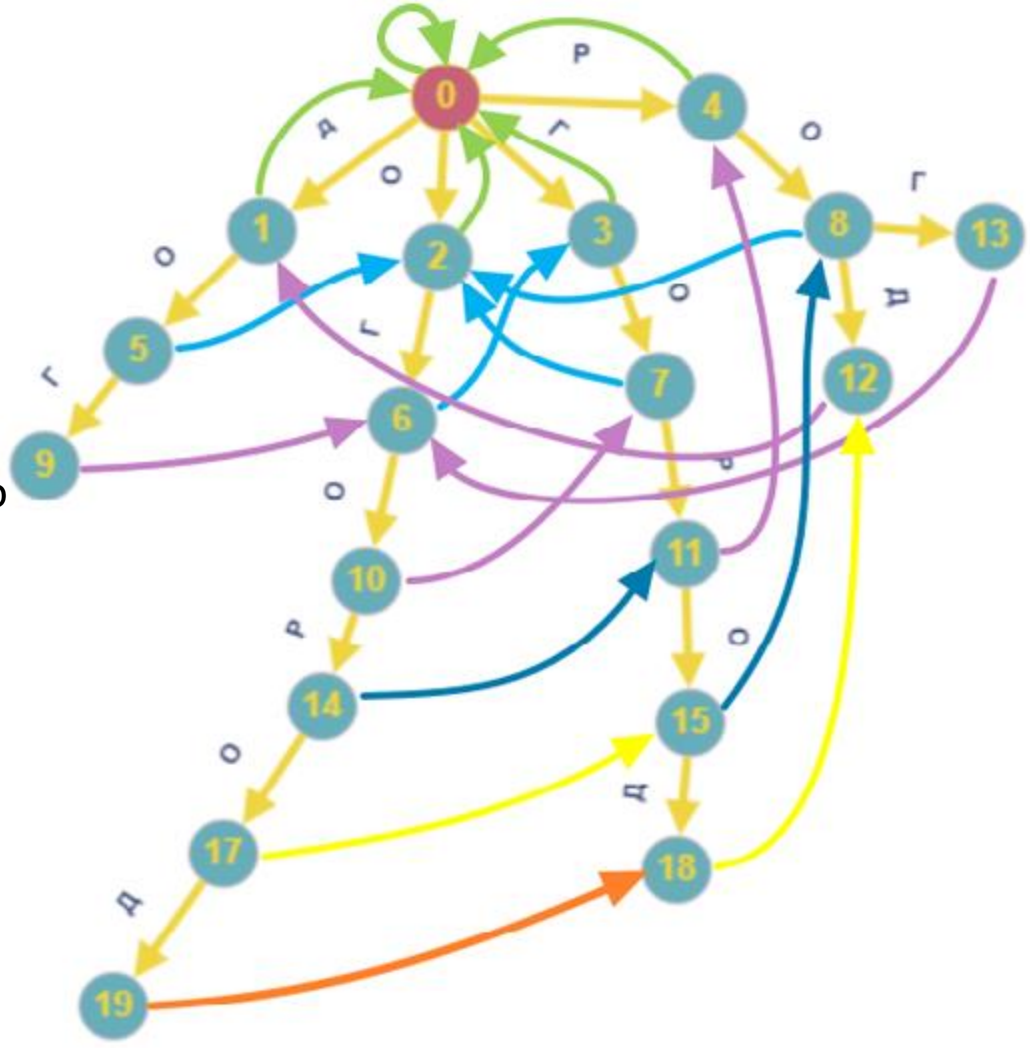


Таблица переходов (полного) автомата Ахо-Корасик

Аналогично получаем переходы в автомате Ахо-Корасика для других вершин.
Имеем следующую **таблицу переходов (для номеров вершин *v*)**

	v/Σ	д	о	г	р	
→	**0**	1	2	3	4	
	1	1	5	3	4	
	2	1	2	6	4	
	3	1	7	3	4	
	4	1	8	3	4	
	5	1	2	9	4	
	6	1	10	3	4	
	7	1	2	6	11	
	8	12	2	13	4	
	9	1	10	3	4	→

v/Σ	д	о	г	р	
10	1	2	6	14	
11	1	15	3	4	
12	1	2	5	4	→
13	1	10	3	4	→
14	1	17	3	4	
15	18	2	13	4	
17	19	2	6	4	
18	1	5	3	4	→
19	1	5	3	4	→

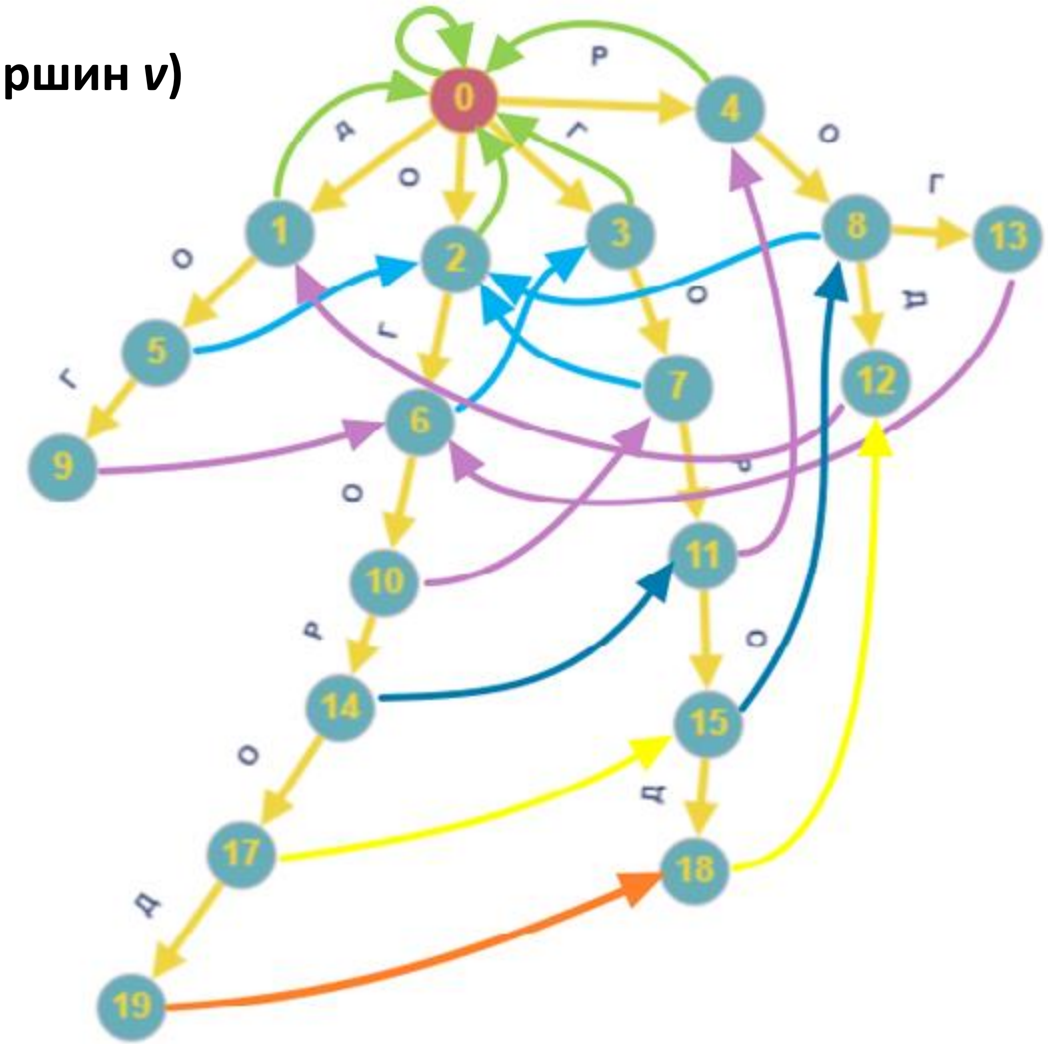


Таблица переходов (полного) автомата Ахо-Корасик

Аналогично получаем переходы в автомате Ахо-Корасика для других вершин.
Имеем следующую **таблицу переходов (для префиксов *v*’)**

v’/Σ	д	о	г	р
λ	д	о	г	р
д	д	до	г	р
о	д	о	ог	р
г	д	го	г	р
р	д	ро	г	р
до	д	о	дог	р
ог	д	ого	г	р
го	д	о	ог	гор
ро	род	о	рог	р
дог	д	ого	г	р

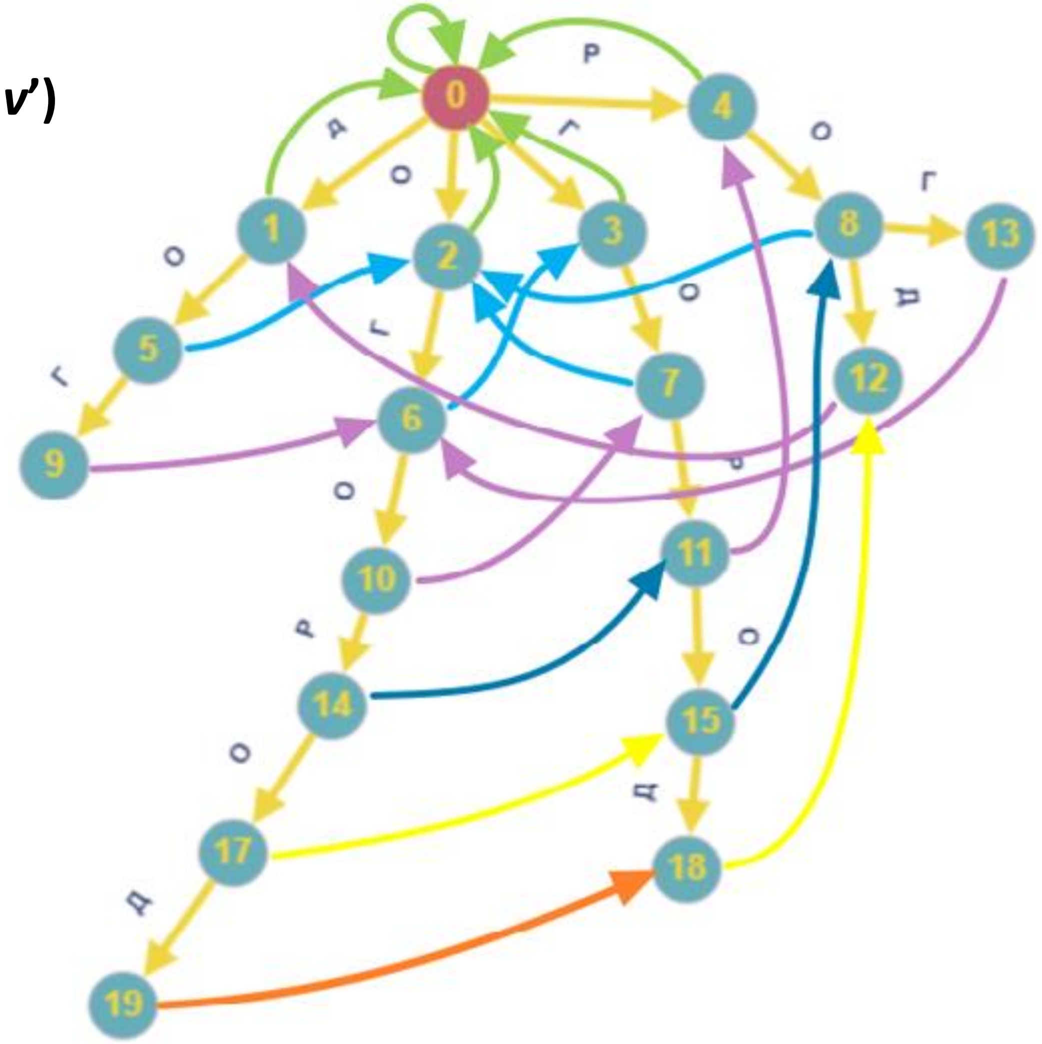
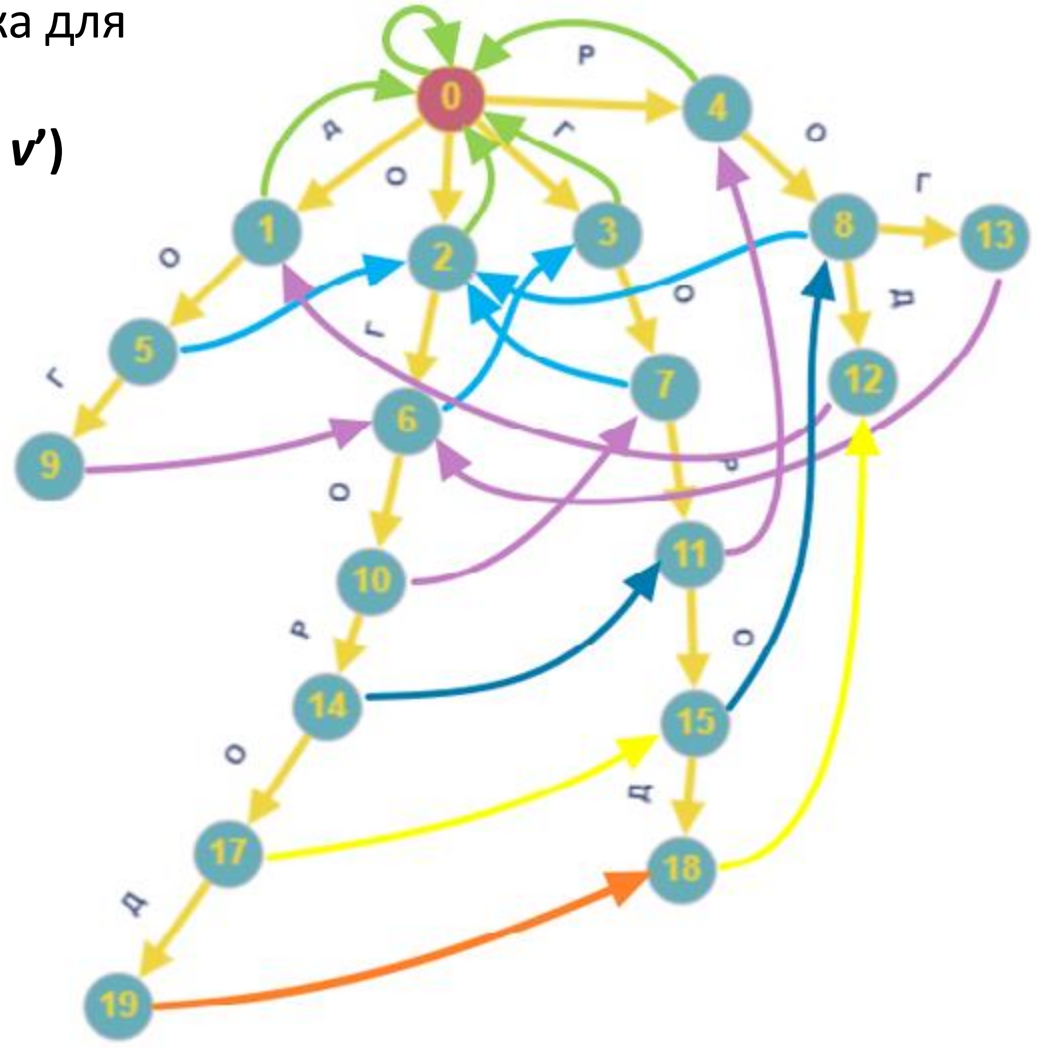


Таблица переходов (полного) автомата Ахо-Корасик (продолжение)

Аналогично получаем переходы в автомате Ахо-Корасика для других вершин.
Имеем следующую **таблицу переходов (для префиксов *v’*)**

v’/Σ	Д	О	Г	Р	
ОГО	Д	О	ОГ	ОГОР	
ГОР	Д	ГОРО	Г	Р	
РОД	Д	ДО	Г	Р	→
РОГ	Д	ОГО	Г	Р	→
ОГОР	Д	ОГОРО	Г	Р	
ГОРО	РОД	О	РОГ	Р	
ОГОРО	ОГОРОД	О	РОГ	Р	
ГОРОД	Д	ДО	Г	Р	→
ОГОРОД	Д	ДО	Г	Р	→

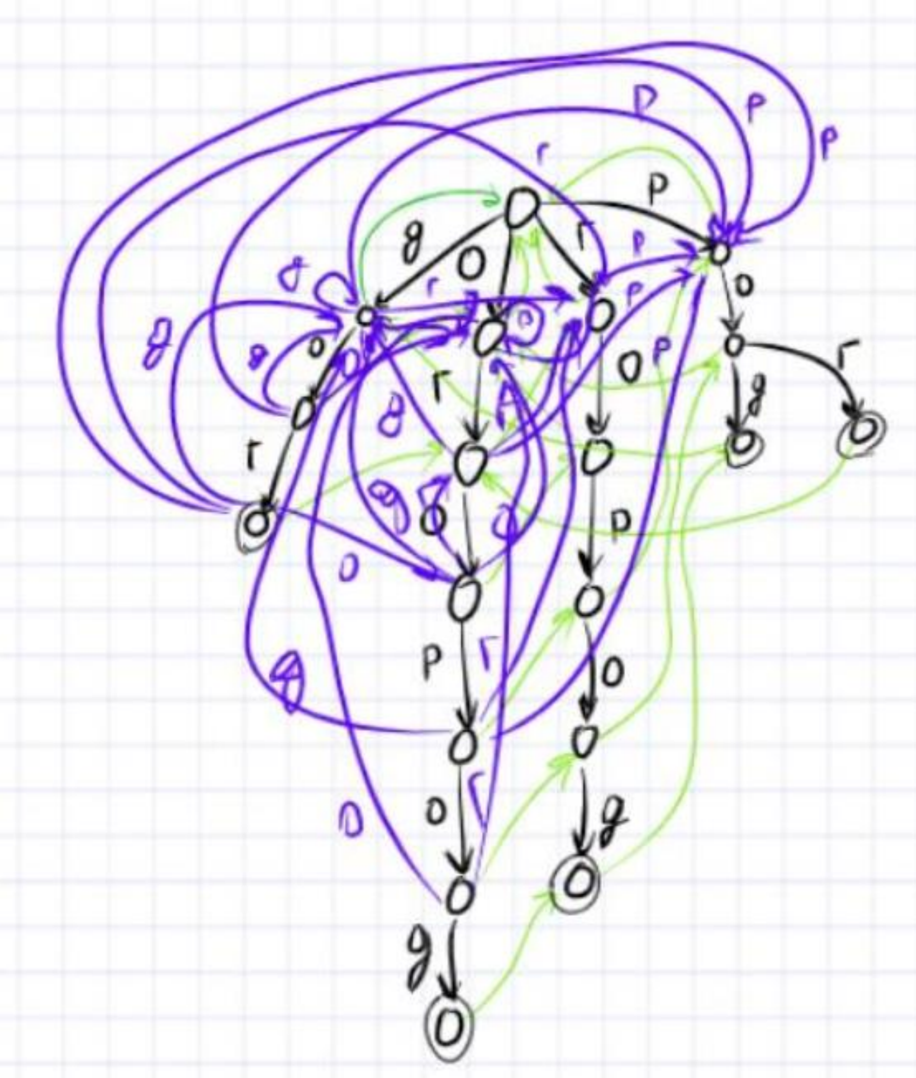


Графическое представление (полного) автомата Ахо-Корасика

Можно, конечно, изобразить (полный) **автомат Ахо-Корасик** в виде **графа**.

Здесь зелеными дугами обозначены суффиксные ссылки, а фиолетовыми дугами – переходы автомата.

Видно, что такой граф автомата трудно воспринимать, а легче работать с **таблицей переходов** этого автомата.



Применение автомата Ахо-Корасик

Автомат Ахо-Корасик используется для нахождения **шаблонных слов** (т.е. слов из словаря) в данном **слове** (т.е. в тексте). **Суффиксная ссылка** называется еще **функцией ошибки (функцией отката, префиксной функцией),** а **бор - префиксным деревом**.

Например, слово **ДОРОГ** содержит шаблонное слово **РОГ** в качестве суффикса, и поэтому принимается автоматом. Действительно, имеем пусть

$$\to \lambda \xrightarrow{Д} Д \xrightarrow{О} ДО \xrightarrow{Р} Р \xrightarrow{О} РО \xrightarrow{Г} РОГ \to$$

А слово **ДОРОГО** автоматом не принимается:

$$\to \lambda \xrightarrow{Д} Д \xrightarrow{О} ДО \xrightarrow{Р}$$

$$Р \xrightarrow{О} РО \xrightarrow{Г} РОГ \xrightarrow{О} О$$

v’/Σ	Д	О	Г	Р
→ λ	Д	О	Г	Р
Д	Д	ДО	Г	Р
О	Д	О	ОГ	Р
Г	Д	ГО	Г	Р
Р	Д	РО	Г	Р
ДО	Д	О	ДОГ	Р
ОГ	Д	ОГО	Г	Р
ГО	Д	О	ОГ	ГОР
РО	РОД	О	РОГ	Р
ДОГ	Д	ОГО	Г	Р →

v’/Σ	Д	О	Г	Р
ОГО	Д	О	ОГ	ОГОР
ГОР	Д	ГОРО	Г	Р
РОД	Д	ДО	Г	Р →
РОГ	Д	ОГО	Г	Р →
ОГОР	Д	ОГОРО	Г	Р
ГОРО	РОД	О	РОГ	Р
ОГОРО	ОГОРОД	О	РОГ	Р
ГОРОД	Д	ДО	Г	Р →
ОГОРОД	Д	ДО	Г	Р →