Лабораторная работа 2
(вариант 1)

1. Исходное регулярное выражение

Алфавит: $\Sigma = \{a, b\}$.

Регулярное выражение:

$$R = \big((a|b)^*\, abb\, (a|b)\big) \mid \big((a|ba)^*\, bb\, (a^*b)^*\, aba\, (a|b)\big).$$

2. НКА

Построенный НКА изображён на рис. 1.

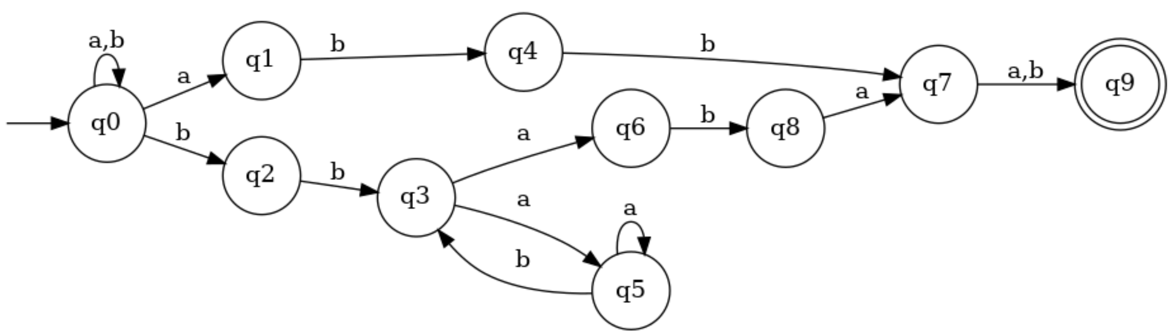

Рис. 1: НКА

2.1. Таблица префиксов и суффиксов НКА

	ε	a	aa	abab	ababaa	abba	ba	baa	babaa	bba
ε	0	0	0	0	0	1	0	0	0	0
a	0	0	0	0	0	1	0	0	0	1
ab	0	0	0	0	0	1	1	0	1	0
abb	0	1	0	1	1	1	0	0	1	0
abba	1	0	0	0	1	1	0	1	1	1
b	0	0	0	0	0	1	0	0	1	0
bb	0	0	0	1	1	1	0	0	1	0
bba	0	0	0	0	1	1	0	1	1	1
bbaa	0	0	0	0	1	1	0	0	1	1
bbab	0	0	1	1	1	1	1	0	1	1

Таблица 1: Таблица префиксов и суффиксов для НКА

3. ДКА и минимизация

На рис. 2 показан минимальный ДКА.

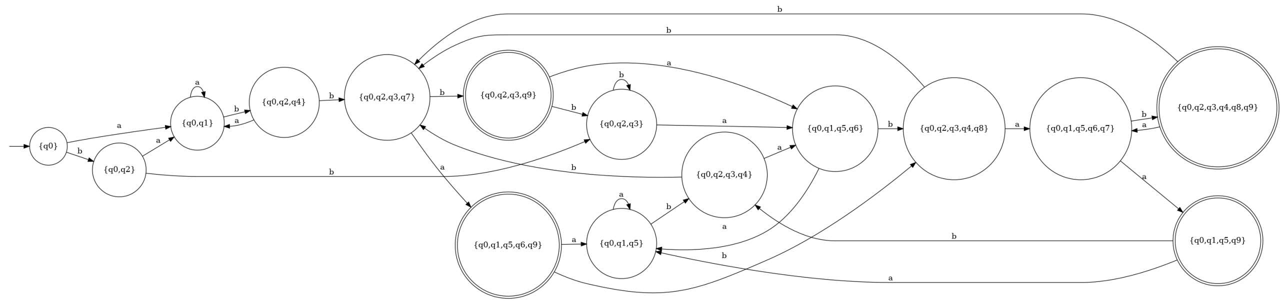

Рис. 2: минимальный ДКА

3.1. Классы эквивалентности

Отметим классы эквивалентности: рис. 3.

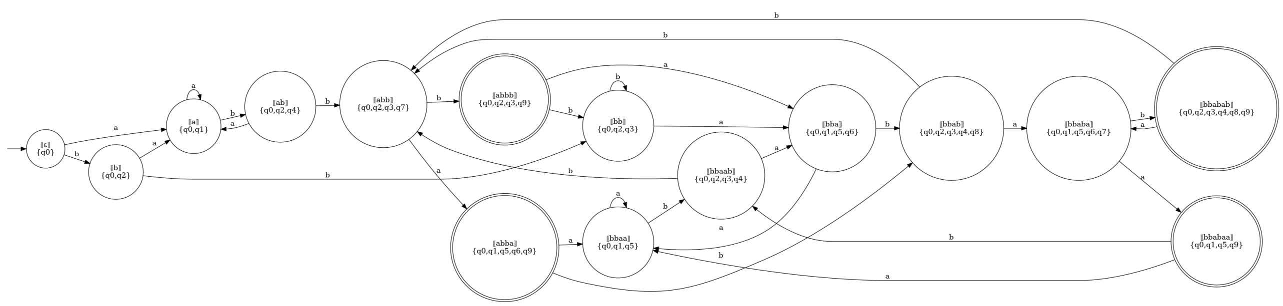

Рис. 3: Минимальный ДКА

3.2. Проверка минимальности

Построим таблицу эквивалентности. Если все строки различны, то состояния попарно различны, значит ДКА минимален.

	ε	a	aa	abaa	abba	ba	baa	babaa	bba
ε	0	0	0	0	1	0	0	0	0
a	0	0	0	0	1	0	0	0	1
b	0	0	0	0	1	0	0	1	0
ab	0	0	0	0	0	1	0	1	0
abb	0	1	0	1	1	0	0	1	0
abba	1	0	0	0	0	0	1	1	1
abbb	1	0	0	1	1	0	0	1	0
bb	0	0	0	1	1	0	0	1	0
bba	0	0	0	0	1	0	0	1	1
bbaa	0	0	0	0	1	0	1	1	1
bbaab	0	0	0	1	1	1	0	1	0
bbab	0	0	1	1	1	1	0	1	0
bbaba	0	1	0	0	1	0	1	1	1
bbabaa	1	0	0	0	1	0	0	1	1
bbabab	1	0	1	1	1	1	1	1	0

Таблица 2: Таблица различимости

Все строки различны, значит он минимален.

4. Расширенное регулярное выражение и ПКА

4.1. Расширенная регулярка

Исходное выражение:

$$R = \big((a|b)^*\, abb\,(a|b)\big) \;\mid\; \big((a|ba)^*\, bb\,(a^*b)^*\, aba\,(a|b)\big).$$

Перепишем его в расширенном виде:

$$R \;=\; \hat{}\Big((a|b)^*\,(? <= abb)\,(a|b) \;\mid\; (a|b)^*\,(? = bb)\,(a|b)^*\,(? <= baba)\,(a|b)\Big)\$.$$

Почему первая ветка эквивалентна $(a|b)^*abb(a|b)$**.** Конструкция $(a|b)^*$ разрешает произвольный префикс над $\{a, b\}$. Далее $abb(a|b)$ означает, что слово заканчивается на $abba$ или на $abbb$.

Это записано как $(a|b)^*(? <= abb)(a|b)$: проверка $(? <= abb)$ требует, чтобы прямо перед последним символом стояло abb, а затем $(a|b)$ задаёт этот последний символ.

Почему вторая ветка эквивалентна $(a|ba)^*\, bb\,(a^*b)^*\, aba\,(a|b)$. Заметим, что как только в слове появляется подслово `bb`, всё, что стоит до этого подслова, перестаёт иметь значение для распознавания: дальше достаточно проверить, что `bb` действительно встречается где-то внутри слова, и что ближе к концу выполняется все остальное.

Поэтому префикс до некоторого вхождения `bb` можно обозначить просто как $(a|b)^*$, а сам факт существования позиции, с которой начинается `bb`, зафиксируем $(? = bb)$.

Часть $(a^*b)^*$ порождает строки, которые либо пусты, либо оканчиваются на b (потому что каждый блок a^*b заканчивается символом b). Значит, перед подсловом `aba` в слове может стоять либо ничего, либо b.

Отсюда следует, что перед последним символом слова должно стоять подслово `baba`. $(a|b)$ задаёт последний символ слова.

Это записано как:

$$(a|b)^*(? = bb)\,(a|b)^*(? <= baba)\,(a|b),$$

где $(? <= baba)$ гарантирует наличие `baba` прямо перед последним символом.

4.2. ПКА

ПКА для того же языка изображён на рис. 4.

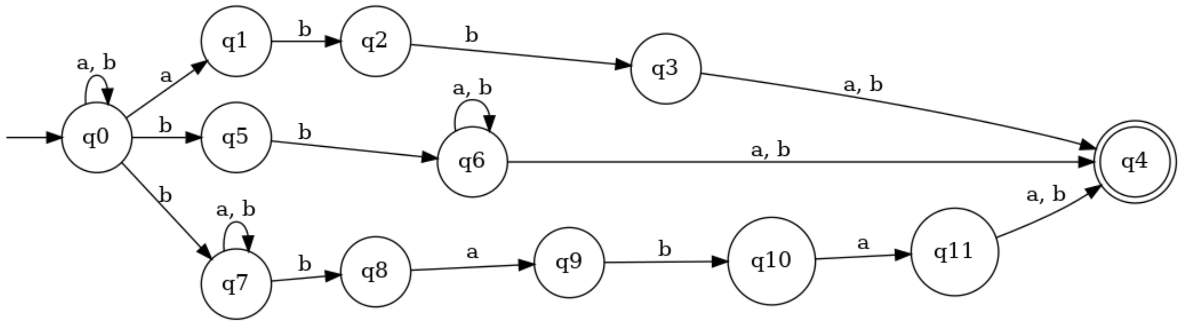

Рис. 4: ПКА

	ε	a	ababaa	baa	babaa	bba
a	0	0	0	0	0	1
ab	0	0	0	0	1	0
abba	1	0	1	1	1	1
bb	0	0	1	0	1	0
bba	0	0	1	1	1	1
bbab	0	0	1	0	1	1

Таблица 3: Таблица префиксов и суффиксов для ПКА