1 Введение

Классическими операциями в теории формальных языков являются объединение и конкатенация. Они могут быть определены в терминах логики следующим образом: объединение $K \cup L$ — это множество строк w таких, что $w \in K$ **или** $w \in L$, которая является дизъюнкцией двух условий; аналогично, строка w содержится в конкатенации $K \cdot L$, если дизъюнкция по всем разбиениям $w = uv$, конъюнкций $u \in K$ *и* $v \in L$.

Заменим операции булевой логики, дизъюнкцию и конъюнкцию, на операции в двухэлементном поле GF(2), *исключающее OR* и конъюнкцию. Таким образом, операция объединения превращается в симметрическую разность, а конкатенация порождает новую операцию — *GF(2)-конкатенацию*, определенную следующим образом:

$$K \odot L = \{\, w \mid \#\ \text{разбиений}\ w = uv,\ \text{где}\ u \in K\ \text{и}\ v \in L,\ \text{нечетно}\,\}$$

2 GF(2)-операции и их основные свойства

По определению, $K \odot L \subseteq K \cdot L$, если конкатенация однозначна, то эти два языка совпадают.

Пример 1. $\{\varepsilon, a\} \odot \{\varepsilon, a\} = \{\varepsilon, aa\}$. *Строка a не принадлежит GF(2)-конкатенации, так как имеет два представления, которые исключают друг друга.*

Пример 2. $a^* \odot a^* = (aa)^*$. *Каждая строка a^n разбивается ровно $n+1$ способами, следовательно, GF(2)-конкатенация содержит только строки четной длины.*

Следующий пример показывает, что некоторые языки имеют *GF(2)-обратный* язык.

Пример 3. $\{\varepsilon, ab\} \odot (ab)^* = \{\varepsilon\}$.

Теорема 1. *Для любого языка $L \subseteq \Sigma^*$ такого, что $\varepsilon \in L$, существует единственный язык $L^{-1} \subseteq \Sigma^*$, который удовлетворяет равенству $L \odot L^{-1} = L^{-1} \odot L = \{\varepsilon\}$.*

Операция обращения может быть эквивалентно представлена следующим видом звездочки Клини.

Определение 1. *Для любого языка L, его GF(2)-звездочка обозначаемая $L^{\circledast}$, это множество всех строк w, имеющих нечетное число представлений в виде $w = w_1 w_2 \dots w_k$, где $k \geqslant 0$ и $w_1, \dots, w_k \in L \setminus \{\varepsilon\}$.*

Теорема 2. *Для любого языка $L \subseteq \Sigma^*$ такого, что $\varepsilon \in L$, обратный язык L^{-1} равен GF(2)-звездочке $L^{\circledast}$.*

3 Замкнутость и сложность GF(2)-операций

Важный вопрос для введенных операций на регулярных языках — сохранение класса регулярных языков и сложность описания этих операций, то есть насколько большой автомат необходим для представления операций на конечных автоматах заданного размера. В этом разделе показано, что GF(2)-конкатация и GF(2)-обратный сохраняют класс регулярных языков, а существуют детерминированные конечные автоматы (DFA), представляющие эти операции, которые являются оптимальными по числу состояний.

3.1 GF(2)-конкатенация

Теорема 3. *Для любых двух DFA $\mathcal{A} = (\Sigma, P, p_0, \eta, E)$ и $\mathcal{B} = (\Sigma, Q, q_0, \delta, F)$, GF(2)-конкатенация $L(\mathcal{A})$ и $L(\mathcal{B})$ распознаётся DFA $\mathcal{C}$ со множеством состояний $P \times 2^Q$.*

Теорема 4. *Для любых $m, n \geqslant 3$ существуют языки K и L над алфавитом $\Sigma = \{a, b\}$, распознаваемые DFA на m и n состояниях, соответственно, для каждого DFA, распознающего их GF(2)-конкатенацию, $K \odot L$ должен иметь по крайней мере $m \cdot 2^n$ состояний.*

3.2 GF(2)-обратный

Теорема 5. *Для всякого DFA $\mathcal{A} = (\Sigma, Q, q_0, \delta, F)$ язык $L(\mathcal{A})^{\circledast}$ распознается некоторым DFA $\mathcal{C}$ со множеством состояний $2^Q \cup \{q_0'\}$.*

Теорема 6. *Для каждого целого $n \geqslant 3$ существует язык L над алфавитом $\Sigma = \{a, b, c\}$, где $\epsilon \in L$, который распознается DFA с n состояниями, такой, что любой DFA, распознающий его GF(2)-обратный, L^{-1} должен иметь по крайней мере $2^n + 1$ состояний.*

4 Односимвольный случай

4.1 GF(2)-конкатенация

Теорема 7. *Пусть $\mathcal{A}$ и $\mathcal{B}$— два DFA над односимвольным алфавитом с m и с n состояниями соответственно. Тогда GF(2)-конкатенация $L(\mathcal{A}) \odot L(\mathcal{B})$ распознается DFA с $2mn$ состояниями.*

4.2 GF(2)-обратный

Теорема 8. *Для всякого DFA* $\mathcal{A} = (\Sigma, Q, q_0, \delta, F)$ *язык* $L(\mathcal{A})^{\circledast}$ *распознается некоторым DFA* $\mathcal{C}$ *с* $2^{n-1}+1$ *достижимым состоянием.*