4. Регулярные выражения

Разделы:

- Построение РВ
- Операторы РВ
- Связь КА и РВ

Регулярные выражения

- РВ определяют те же языки, что и различные типы КА, а именно регулярные языки
- В отличие от автоматов РВ позволяют определять допустимые строки декларативным способом
- Поэтому РВ используются в качестве входного языка во многих системах, обрабатывающих цепочки:
  - grep/egrep
  - flex
- РВ 01*+10* определяет язык всех строк, состоящих из одного нуля, за которым следует произвольное количество 1, либо из одной единицы, за которой следует произвольное количество 0

Операторы РВ

- **Объединение** языков L и M – это множество строк, которые содержатся либо в L, либо в M, либо в обоих языках.
  - если $L = \{001, 0, 11\}$ и $M = \{\varepsilon, 11\}$, то их объединение дает $\{\varepsilon, 11, 001, 0\}$
- **Конкатенация** языков L и M – это множество строк, образуемых дописыванием к любой строке из L любой строки из M
  - Если $L = \{001, 0, 11\}$ и $\mathrm{M} = \{\varepsilon, 11\}$, то их конкатенация $LM = \{001, 00111, 0, 011, 11, 1111\}$
- **Итерация** языка L обозначается как L^* и представляет собой множество всех тех строк, которые можно образовать конкатенацией любого количества строк из L
  - Если $L = \{0, 1\}$ то L^* образуют все строки из 0 и 1

Построение РВ

- РВ можно определить рекурсивно, не только характеризуя правильные РВ, но и для каждого РВ E описывая представленный ими язык $L(E)$
  - Базис:
    - Константы ε и пустое множество являются РВ, определяющими языки $\{\varepsilon\}$ и пустое множество соответственно
    - Если a – произвольный символ, то a – РВ, определяющее язык $\{a\}$
    - Переменная, записываемая заглавным курсивным символом, представляет собой произвольный язык
  - Индуктивный шаг:
    - Если E и F – РВ, то $E + F$ – РВ, определяющее объединение языков
    - Если E и F – РВ, то EF – РВ, определяющее конкатенацию языков
    - Если E – РВ, то E^* – РВ, определяющее итерацию языка $L(E)$
    - Если E – РВ, то (E) – РВ, определяющее тот же язык $L(E)$, что и выражение E.

Построение РВ

От РВ к НКА

Шаг 1. Стартуем от РВ R.

Шаг 2. Создаем обобщенный граф переходов G с единственным начальным состоянием q_0, единственным конечным состоянием q_1 и единственным переходом между ними, помеченным исходным РВ R.

Шаг 3. Хотя существует некоторый переход t, принадлежащий G, из состояния q_i в состояние q_j, помеченный выражением S, состоящим более, чем из одного символа, пусть φ – это оператор верхнего уровня для выражения S, и пусть $[\alpha_1, \alpha_2, ..., \alpha_\psi]$ – это упорядоченный список операндов оператора φ (т.к. скобки и звездочка Клини имеют один операнд, то в этом случае $\psi = 1$).

а) Если φ – это круглые скобки, заменяем t на РВ-переход по α_1 из состояния q_i в состояние q_j.

б) Если φ – это оператор итерации, то создаем в G два новых состояния q_x и q_y, а также переход по α_1 между ними, и затем четыре ε-перехода: из q_i в состояние q_x, из q_y в q_j, из q_i в q_j, из q_j в q_i.

в) Если φ – это оператор объединения, то удаляем t, и для всех k от 1 до ψ выполняем следующее: создаем два новых состояния q_{xk} и q_{yk}; создаем переход по α_k между q_{xk} и q_{yk}; затем два ε-перехода: из q_i в состояние q_{xk}, из q_{yk} в q_j.

г) Если φ – это оператор конкатенации, то удаляем t, и для всех k от 1 до ψ выполняем следующее: создаем два новых состояния q_{xk} и q_{yk}; создаем переход по α_k между q_{xk} и q_{yk}; если $\psi > 0$ создаем ε-переход из q_{yk-1} в состояние q_{xk}. Наконец, создаем два ε-перехода: из q_i в состояние q_{x1}, из q_ψ в q_j.

Шаг 4. Мы получаем GTG, являющийся правильным НКА.

От НКА к РВ

Шаг 1. Стартуем от КА, который рассматриваем как обобщенный граф G.

Шаг 2. Пусть F – это множество заключительных состояний G, а q_0 – начальное состояние. Если $|F| > 1$ или $F = \{ q_0\}$, то создаем новое состояние q_f , производим все ε-переходы для каждого q_i из F от q_i к q_f, и делаем q_f – единственным конечным состоянием.

Шаг 3. Пусть S – это множество всех состояний G. Для каждой пары (q_i, q_j) из SxS пусть $L = \{l_1, l_2, ..., l_n\}$ будет множеством всех РВ на переходах от q_i к q_j. Пусть $e = \varnothing$, если $|L| = 0$, и $e = l_1 + l_2 +... + l_n$, в противном случае. Заменяем все переходы от q_i к q_j с единственным переходом между ними по выражению e.

Шаг 4. Пусть T – это множество всех нестартовых и незаключительных состояний G. Пусть r_{xy} – это выражение на переходе от q_x к q_y. Для каждого q_k из T и каждой пары (q_i, q_j) из (T-$\{q_k\}$)x(T-$\{q_k\}$) заменяем выражение r_{ij} на $r_{ij} + r_{ik}r_{kk}^*r_{kj}$ и удаляем q_k из G.

Шаг 5. У G теперь два требуемых состояния, а эквивалентное РВ теперь $r = (r_{00}^* r_{0f} r_{ffj}^* r_{f0})^* r_{00}^* r_{0f} r_{ff}^*$.

Дополнительные источники

- Гилл, А. Введение в теорию конечных автоматов / А. Гилл. – М.: Наука, 1966. – 272 с.
- Кузнецов, А.С. Теория вычислительных процессов [Текст] : учеб. пособие / А. С. Кузнецов, М. А. Русаков, Р. Ю. Царев ; Сиб. федерал. ун-т. - Красноярск: ИПК СФУ, 2008. – 184 с.
- Короткова, М.А. Математическая теория автоматов. Учебное пособие / М.А. Короткова. – М.: МИФИ, 2008. – 116 с.
- Молчанов, А. Ю. Системное программное обеспечение. 3-е изд. / А.Ю. Молчанов. – СПб.: Питер, 2010. – 400 с.

Дополнительные источники

- Теория автоматов / Э. А. Якубайтис, В. О. Васюкевич, А. Ю. Гобземис, Н. Е. Зазнова, А. А. Курмит, А. А. Лоренц, А. Ф. Петренко, В. П. Чапенко // Теория вероятностей. Математическая статистика. Теоретическая кибернетика. — М.: ВИНИТИ, 1976. — Т. 13. — С. 109–188. — URL http://www.mathnet.ru/php/getFT.phtml?jrnid=intv&paperid=28&what=fullt&option_lang=rus
- Серебряков В. А., Галочкин М. П., Гончар Д. Р., Фуругян М. Г. Теория и реализация языков программирования — М.: МЗ-Пресс, 2006 г., 2-е изд. - http://trpl7.ru/t-books/TRYAP_BOOK_Details.htm
- Введение в схемы, автоматы и алгоритмы - http://www.intuit.ru/studies/courses/1030/205/info