Элементы дискретной математики в статическом анализе кода

Статический анализ кода - это анализ исходного кода программ без их выполнения. Здесь упомянуты те разделы дискретной математики и теоретической информатики, которые наиболее часто возникают в задаче стат. анализа.

Напомним

Определение. Язык L - множество строк w над некоторым алфавитом Σ.

Определение. Машиной Тьюринга называется семерка $(Q, \Gamma, \Sigma, \delta, B, q_0, F)$, где
Q - конечное непустое множество состояний,
Γ - конечное непустое множество символов алфавита,
B - пустой символ,
$\Sigma \subset \Gamma \setminus B$ - множество входных символов,
$q_0 \in Q$ - начальное состояние,
$F \subset Q$ - множество финальных состояний,
$\delta : Q \setminus F \times \Gamma \to Q \times \Gamma \times \{L, R, \bullet\}$ - функция перехода.

Обозначим через $L(M)$ множество всех строк, на которых машина останавливается в финальном состоянии.

Определение. Язык L является рекурсивно-перечислимым (РП), если $L = L(M)$ для некоторой машины Тьюринга M. Иначе говоря, L допустим с помощью некоторой машины Тьюринга.

Определение. Множество натуральных чисел X *разрешимо*, если существует алгоритм, который по любому натуральному n определяет, принадлежит ли оно множеству X. Другими словами, множество X разрешимо, если его характеристическая функция $\chi(n) = (\text{if } n \in X \text{ then } 1 \text{ else } 0)$ вычислима. В противном случае множество неразрешимо. Аналогично определяются другие разрешимые множества.

Определение. *Свойство* РП-языков представляет собой некоторое множество РП-языков. Свойство называется *тривиальным*, если оно либо пустое, либо содержит все РП-языки. Иначе свойство называется *нетривиальным*.

Свойство *разрешимо*, если соответствующее множество языков разрешимо.

Теперь приведем значимое утверждение:

Теорема. *(Райс): всякое нетривиальное свойство рекурсивно-перечислимых языков неразрешимо.*

Откуда следует, что в общем виде задача статического анализа неразрешима. Поэтому задача решается приближенно, в зависимости от цели жертвуя надежностью (soundness) или полнотой (completeness).

Кроме перечисленного выше, в стат. анализе используются формальные методы (например, абстрактная интерпретация и символьное выполнение), графы (потоки управления), в частности деревья (абстрактные синтаксические деревья, AST), регулярные выражения, конечные автоматы, формальные грамматики, другие модели вычислений (потоки вычисления), математическая логика (построение булевых уравнений и их решение при помощи SMT-солверов).

Зная теорию для используемых математических объектов, можно утверждать о корректности нашей или стороней программы.