Автоматні граматики та мови

Розглянемо автоматні граматики та мови докладніше, маючи на меті побудову алгоритмів розпізнавання цього класу мов.

Граф автоматної граматики

Для кожної автоматної граматики можна побудувати спрямований граф за такими правилами:

1. Кожному нетермінальному символу граматики ставиться у відповідність вершина графа, що позначається цим символом.

2. За наявності правил виду

А →a

додається додаткова вершина, яка позначається символом К.

3. Кожне правило виду

А→аВ

породжує дугу графа, що веде з вершини А у вершину В.



Дуга позначається символом а

4. Кожне правило виду

А →a

породжує дугу графа, яка веде з вершини А у вершину К



Дуга позначається символом а

5. Вершина, що відповідає початковому нетерміналу, позначається стрілкою



6. Вершина К та вершини, що відповідають нетерміналам, для яких є правило

А →ε

позначаються як кінцеві. Ми зображатимемо їх подвійним гуртком.



Побудуємо граф автоматної граматики G10 (рис. 9). Двом нетерміналам цієї граматики відповідатимуть вершини N та В (п. 1). Оскільки в граматиці є кілька правил, у правій частині яких записано єдиний термінал, додамо вершину К (п. 2).



Рис. 9. Граф автоматної граматики G10

З'єднаємо вершини дугами, як це наказується п. 3 і п. 4.

Вершину N позначимо стрілкою як початкову (п. 5).

Граф автоматної граматики може використовуватися для породження ланцюжків мови. Будь-який шлях з початкової вершини графа в одну з кінцевих вершин породжує ланцюжок терміналів, що відповідає прохідним дугам. Цей ланцюжок належить мові, що народжується граматикою. І, навпаки, для будь-якої сентенції граматики можна знайти шлях, що веде з початкової вершини в одну з кінцевих і проходить по дугах, позначених символами цієї сентенції.

Граматика G10 породжує мову ідентифікаторів. Неважко переконатися, що для будь-якого ідентифікатора знайдеться шлях з вершини N до вершини К, а будь-який шлях з N відповідає правильному ідентифікатору. Граф автоматної граматики ідентичний діаграмі переходів кінцевого автомата - абстрактного пристрою, що є моделлю певного класу реальних автоматичних пристроїв та об'єктом вивчення теорії автоматів.

**Скінчені автомати**

Скінченим автоматом (СА) називається п'ятірка:

A=(N,T,P,S,F),

де N - кінцева множина станів автомата;

Т-вхідний алфавіт - кінцева множина символів;

Р-функція переходів автомата (загалом неоднозначна), що відображає безліч пар стан - вхідний символ у множину станів;

S - Початковий стан. S∈N;

F-множина кінцевих (фінітних) станів. F⊆N.

Скінчений автомат діє в такий спосіб. Спочатку він перебуває у стані S. На вхід СА надходять символи, що належать до вхідного алфавіту. Послідовність вхідних символів утворює вхідний ланцюжок. Перебуваючи в деякому стані і отримавши на вхід черговий символ, автомат переходить у наступний стан, що визначається значенням функції переходів для цієї пари символ-стан, і зчитує черговий символ.

У загальному випадку функція переходів може визначати перехід у кілька станів для цієї пари символ-стан. У цьому випадку говорять про недетермінований кінцевий автомат (НСА). Автомат зупиняється, коли символи закінчуються на його вході. Якщо, прочитавши вхідний ланцюжок а, автомат зупинився у певному стані, кажуть, що ланцюжок а перевела автомат із початкового стану у стан В. Якщо В — одне з кінцевих станів (В∈F), то кажуть, що автомат приймає (допускає) ланцюжок а.

*Безліч всіх ланцюжків, що переводять скінчений автомат А з початкового в один із кінцевих станів (множина ланцюжків, що приймаються СА), утворює мову L(A), що приймається (допускається) СА.*

*Мова, що породжується автоматною граматикою G, збігається з мовою, яка приймається відповідним скінченим автоматом А.*

*L(G) = L(A).*

Як ми бачили, СА може задаватися за допомогою діаграми переходів. Наприклад, граф автоматної граматики G10 показаний на рис. 9 може вважатися діаграмою переходів автомата A10. При цьому

L(G10) = L(A10).

**Перетворення недетермінованого скінченого автомата на детермінований скінчений автомат**

Та обставина, що з переходом від автоматної граматики до СА ми отримуємо у випадку недетермінований скінчений автомат (НСА), утрудняє його використання ролі розпізнавача автоматної мови.

 Недетермінованість автомата виявляється у тому, що з деяких вершин його діаграми переходів є кілька дуг, що виходять з цих вершин і позначені одним і тим самим символом.

Так, автомат, зображений на рис. 9 є недетермінованим.

З вершини N виходять дві дуги, позначені символом **а**, з вершини В — по дві дуги, позначені символами **а** і **b**.

Було б дуже бажано мати можливість будувати для автоматної граматики детермінований скінчений автомат (ДСА).

***Теорема Кліні****. Для кожного НСА можна побудувати ДСА, що допускає ту саму мову*

Розглянемо алгоритм побудови ДСА, еквівалентного даному НСА.

Для ілюстрації алгоритму будемо застосовувати його до НСА A10 (див. рис. 9), який приймає мову ідентифікаторів.

 1. Нехай вихідний НСА має k станів. Для побудови ДCА візьмемо 2k - 1 станів, кожен з яких відповідає одному елементу множини всіх підмножин станів вихідного автомата, крім пустої множини.

Автомат А10 має три (k = 3) стани:

N, B и К.

Новий автомат матиме 23 -1= 7 станів. Вони відповідають таким множинам станів вихідного НCА:

{N}, {В}, {К}, {N, В}, {В, К}, {N, К), {N, В, К}.

Будемо позначати стани нового автомата просто послідовностями літер:

N, В, К, NB, ВК, NK, NBK.

2. З кожного стану S нового автомата направимо не більше ніж один перехід, позначений цим символом, в такий стан, який відповідає множині станів НСА, в які є переходи по цьому символу хоча б з одного стану НСА, що утворює S.

У НСА A10 зі стану N є перехід за символом **а** в стани В і К (див. рис. 9). Отже, зі стану N нового автомата дугу, позначену символом **а**, направляємо в стан ВК.

Розглядаючи стан NB нового автомата, з'ясовуємо, що переходи зі стану N за символом **а** у вихідному автоматі є в стані В і К, зі стану В вихідного автомата - також в стани В і К.

Направляємо дугу, помічену **а**, зі стану NB в стан ВК (рис. 10). Переходу за символом **b** зі стану N у вихідному автоматі немає. Зі стану вихідного автомата є переходи, помічені **b**, у стани В і К. Направляємо дугу **b** зі стану NB в стан ВК.



Рис.10, Детермінований скінчений автомат А11, еквівалентний недетермінованому автомату А10

Аналогічно формуємо переходи з інших станів нового автомата А11, який буде еквівалентний A10. У прикладі виявляється, що усі дуги, позначені як символом **а**, так і символом **b**, ведуть у стан ВК.

3. Як початкове відзначимо стан ДСА, що має те ж позначення, що і початковий стан вихідного НСА. Як кінцеві відзначимо всі стани ДСА, в які входить хоча б один із кінцевих станів вихідного НСА. У прикладі початковим буде стан N Кінцевими станами ДСА будуть всі стани, які включають стан К вихідного автомата, тобто

К, ВК, NK,NBK (см. рис. 10).

Автомат, що вийшов, є детермінованим (з будь-якого стану виходить не більше однієї дуги, позначеної цим символом) і приймає ту ж мову, що і вихідний недетермінований автомат.

Детермінований автомат А11 має більше станів ніж вихідний НСА А10. Неважко, однак, побачити можливості спрощення ДСА, що вийшов.

Більшість його станів (B, K, NB, NK, NBK) недосяжні з початкового стану, тому можуть бути відкинуті.

ДСА А12, що виходить після цього, показаний на рис. 11.

Позначення стану ВК спрощено, воно знову названо просто В. Цей автомат не тільки детермінований, а й простіше за вихідний НСА A10.



Рис. 11. Мінімальний детермінований кінцевий автомат A12, що розпізнає ідентифікатори

По діаграмі переходів автомата, що вийшов, можна знову записати автоматну граматику, що породжує мову ідентифікаторів, еквівалентну граматиці G10, але містить менше правил:

G12: N→aB

 B→aB|bB|**ε**.

До речі, сенс нетерміналу В у новій граматиці зберігся — це «хвіст», який завершує ідентифікатор.

Детермінований скінчений автомат можна розглядати як розпізнавач автоматної мови — пристрій, з допомогою якого легко й ефективно вирішується завдання розбору для автоматної граматики.

У зв'язку з цим поряд з автоматами, які приймають (допускають) деяку мову, говоритимемо про автомати, що розпізнають мову.

Для будь-якої автоматної мови можна побудувати скінчений детермінований автомат, що розпізнає цю мову.

Завдання отримання більш простого ДСА також має загальне рішення.

Для будь-якої автоматної мови можна побудувати єдиний ДСА, який розпізнає цю мову і має мінімально можливу кількість станів.