

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

УТВЕРЖДАЮ
Проректор по учебной и методической работе
_____ Д.А. Зубцов
20 августа 2016 г.

П Р О Г Р А М М А

по дисциплине: **Теория и реализация языков
программирования**

по направлению: 03.03.01 «Прикладные математика и физика»

факультет: **ФУПМ**

кафедра: **математических основ управления**

курс: 2

семестр: 3

Трудоёмкость: вариативная часть – 3 зач. ед.

лекции – 30 часов

Экзамен – 3 семестр

практические (семинарские)

занятия – 30 часов

Диф. зачет – нет

лабораторные занятия – нет

Самостоятельная работа – 18 час.

ВСЕГО АУДИТОРНЫХ ЧАСОВ – 60

Программу составили: д.ф.-м.н. В. А. Серебряков,
к.т.н. Д. Р. Гончар, ст. преп. А. А. Рубцов, к.ф.-м.н. С. П. Тарасов,
ст. преп. К. Б. Теймуразов.

Программа принята на заседании кафедры
математических основ управления
24 апреля 2016 года

Заведующий кафедрой

С. А. Гуз

1. Известные и перспективные направления эффективного применения теории формальных языков как математической дисциплины. Алфавиты, цепочки, языки и их представление. Формальное определение грамматики. Типы грамматик по Хомскому и их свойства. Связь машин Тьюринга и грамматик типа 0. Линейно-ограниченные автоматы и их связь с КЗ-грамматиками.

2. Лексический анализ. Регулярные языки (РЯ) и регулярные выражения (РВ). Конечные автоматы (КА). Детерминированные и недетерминированные КА (ДКА и НКА). Эквивалентность классов языков, определяемых КА, РВ и автоматами грамматики (грамматики типа 3: леволinéйные – ЛЛ, праволinéйные – ПЛ). Свойства замкнутости РЯ. Лемма о накачке для РЯ. Теорема Майхилла-Нероуда и минимальные автоматы. Алгоритмы поиска подстрок и КА. Алгоритм Кнута–Мориса–Пратта (КМП-алгоритм). Линейность алгоритма КМП.

Алгоритмы по теме КА

- Построение ДКА по НКА.
- Построение НКА по РВ.
- Построение ДКА по РВ.
- Построение РВ по НКА.
- По РВ R проверить, что слово принадлежит $L(R)$.
- Построить по языку L язык L^R .
- Построение произведения (конкатенации) РЯ, дополнение РЯ, пересечение РЯ.
- Построение минимального автомата по ДКА.
- КМП-алгоритм.
- Построение по НКА эквивалентных ЛЛ- и ПЛ-грамматик.
- Построение эквивалентного НКА по ЛЛ- и ПЛ-грамматике.
- Решение уравнений с регулярными коэффициентами.

3. Синтаксический анализ: КС-грамматики (КСГ). Преобразования КС-грамматик, приведённые грамматики. Канонические формы КС-грамматик (нормальная форма Хомского). Свойства замкнутости КС-языков (КСЯ), незамкнутость КСЯ относительно пересечения. Дерево вывода КСГ. Однозначность КС-грамматик. Однозначность праволинейной грамматики, построенной по ДКА. Лемма о накачке для КСЯ. Проверка принадлежности слова КСЯ КСГ (алгоритм Кока–Янгера–Касами).

4. Синтаксический анализ: автоматы с магазинной памятью (МА). Детерминированные и недетерминированные МА. Обобщенные МА и их эквивалентность стандартным МА. Эквивалентность МА, распознающих по конечному состоянию (F-МА) и по опустошению магазина (N-МА). Эквивалентность КСГ и МА. Однозначность КСГ, построенной по детерминированному N-МА (без доказательства).

Алгоритмы по теме КСГ и МА

- Удаление недостижимых и бесполезных символов в КСГ. Удаление циклов.
- Удаление левой рекурсии в КСГ.
- Приведение КСГ к нормальной форме Хомского и нормальной форме Грейбах.
- Проверка принадлежности слова КСГ (алгоритм Кока–Янгера–Касами).
- Преобразование N-МА \rightarrow F-МА.
- Преобразование F-МА \rightarrow N-МА.
- Преобразование КСГ в эквивалентный N-МА.
- Преобразование N-МА в эквивалентную КСГ (с доказательством корректности для N-МА с одним состоянием).

5. Дополнительные сведения из теории конечных автоматов. Минимизация числа состояний и эквивалентность детерминированного конечного автомата (ДКА).

6. Предсказывающий разбор *сверху вниз*. Алгоритм разбора *сверху вниз*. Функции *FIRST* и *FOLLOW*. Конструирование таблицы предсказывающего анализатора. LL(1)-грамматики. Удаление левой рекурсии. Левая факторизация. Рекурсивный спуск. LL(k)-грамматики. Разбор *снизу вверх* типа сдвиг-свёртка. Основа. LR(1)-анализаторы. Конструирование LR(1)-таблицы. LR(1)-грамматики. Варианты LR-анализаторов. LR(k)-грамматики.

7. Элементы теории перевода. Синтаксически управляемый перевод. Атрибутные грамматики.

Литература

1. Ахо А., Сети Р., Ульман Дж. Компиляторы. Принципы, технологии, инструменты. М., СПб., Киев: Вильямс, 2001.
2. Мартыненко Б.К. Языки и трансляции. СПб.: СПбГУ, 2004. Доступно по ссылке http://trpl7.ru/t-books/Martin/Martinenko_FLT_Cont.htm
3. Серебряков В. А., Галочкин М. П., Гончар Д. Р., Фуругян М. Г. Теория и реализация языков программирования: учебное пособие для студентов. М.: МЗ-Пресс, 2006. 352 с.
4. Хопкрофт Дж., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. М.: Вильямс, 2002.
5. Ахо А., Лам М., Сети Р., Ульман Дж. Компиляторы. Принципы, технологии и инструментарий. М., СПб., Киев: Вильямс, 2011. 1184 с.

Задание

Задачи, выделенные в дополнительный раздел, а также задачи, помеченные звёздочкой, являются дополнительными и необязательными. Контрольные вопросы являются полноценными задачами, хотя и выделены в отдельные блоки. Решение всех задач должно быть обосновано. Отдельные указания по необходимости обоснования в некоторых задачах являются акцентированием и вовсе не означают, что в других задачах обоснование не требуется. Использование алгоритмов из курса (см. программу), считается обоснованием. При использовании алгоритма проверяющий должен иметь возможность проверить корректность протокола: решения в духе «автомат построен по алгоритму, но вот только ответ» не оцениваются.

Всё вышесказанное относится ко всем письменным работам курса.

Регулярные языки

Задача 1. Определим язык $L \subseteq \{a, b\}^*$ индуктивными правилами:

- 1) $\varepsilon, b, bb \in L$;
- 2) вместе с любым словом $x \in L$ в L также входят слова $ax, bax, bba x$;
- 3) никаких других слов в L нет.

Язык $T \subseteq \{a, b\}^*$ состоит из всех слов, в которых нет трёх букв b подряд.

1. Докажите или опровергните, что $L = T$.¹
2. Запишите язык T в виде регулярного выражения.
3. Постройте конечный автомат, принимающий T . Докажите (по индукции), что построенный автомат принимает язык T .

¹Если равенство неверно, то нужно явно указать слово, принадлежащее одному языку и не принадлежащее другому. Если равенство верно, то нужно провести доказательство по индукции в обе стороны: $L \subseteq T$ и $T \subseteq L$.

Задача 2. Верно ли, что

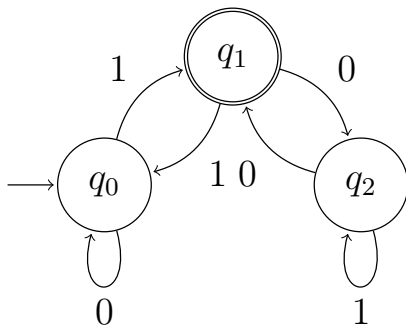
- 1) $\varepsilon \in \{a, aab, aba\}$?
- 2) $\emptyset \in \{a, aab, aba\}$?

Задача 3. Запишите регулярные выражения для языков:

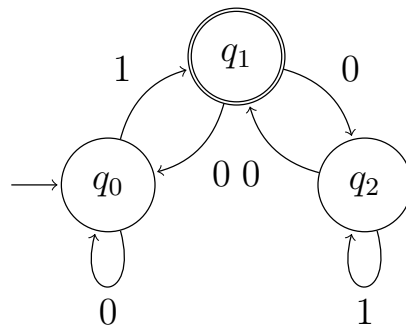
- 1) $\{a^n \mid n > 0\} \times \{b^n \mid n \geq 0\}$;
- 2) $\{a^{3n} \mid n > 0\} \cap \{a^{5n+1} \mid n \geq 0\}^*$.

Задача 4. Автоматы \mathcal{A} и \mathcal{B} заданы диаграммами. Выполните следующие задания.

Автомат \mathcal{A} :



Автомат \mathcal{B} :



Для каждого автомата ответьте на следующие вопросы (1–2).

1. Автомат задан через граф переходов. Запишите определение автомата в виде $(Q, \Sigma, \delta, q_0, F)$. Опишите элементы каждого множества.
2. Является ли автомат детерминированным?

Ответьте на вопросы.

3. Опишите последовательность конфигураций автомата \mathcal{A} при обработке слова $w = 011001$. Верно ли, что $w \in L(\mathcal{A})$?
4. Принимает ли автомат \mathcal{B} слово $v = 0101001$?
5. Укажите по одному слову, принадлежащему $L(\mathcal{A})$, $L(\mathcal{B})$ и по одному слову, не принадлежащему $L(\mathcal{A})$, $L(\mathcal{B})$. Все 4 слова должны быть различными.

Задача 5. Выполните следующие задания.

1. Построить ДКА, принимающий язык L , состоящий из всех слов в алфавите $\{0, 1\}$, которые содержат нечётное число нулей и чётное число единиц.
2. Построить эквивалентную левостороннюю грамматику. Будет ли она однозначной?
3. Построить регулярное выражение для языка L^R .

Задача 6. Будут ли регулярными следующие языки?

1. $L = \{a^{2016n+5} \mid n = 0, 1, \dots\} \cap \{a^{503k+29} \mid k = 401, 402, \dots\} \subseteq \{a^*\}$.
2. $L_2 = \{a^{200n^2+1} \mid n = 1000, 1001, \dots\} \subseteq \{a^*\}$.
3. Язык L_3 всех слов в алфавите $\{0, 1\}$, которые представляют числа в двоичной записи, дающие остаток два при делении на три (слово читается со старших разрядов). Например, 001010 ($1010_2 = 10_{10} = 3 \times 3 + 1$) $\notin L_3$, а 10001 ($10001_2 = 17_{10} = 5 \times 3 + 2$) $\in L_3$.

Задача 7. Постройте НКА, принимающий язык $L_3 = \{\text{Множество слов в алфавите } \{a, b\}, \text{ у которых третий от конца}^2 \text{ символ равен «a»}\}$. Затем, используя алгоритм, построите соответствующий полный ДКА.

Задача 8. Порождает ли регулярное выражение $(ab)^*(ba)^*$ тот же язык, что распознаёт ДКА $M = (\{A, B, C, D\}, \{a, b\}, \delta, A, \{A, D, E\})$, где функция переходов задана следующим образом:

$$\delta(A, a) = B, \delta(A, b) = C, \delta(B, b) = D, \delta(C, a) = E,$$

$$\delta(D, a) = B, \delta(D, b) = C, \delta(E, b) = C.$$

²Последний символ слова равен первому символу с конца слова.

Задача 9. Покажите, что следующий язык удовлетворяет лемме о разрастании для регулярных языков, но сам регулярным не является:

$$L = \{a^k b^{2^i} \mid i, k \geq 0\} \cup \{b^j \mid j = 0, 1, \dots\}.$$

Задача 10. Решите уравнения с регулярными коэффициентами. В каждом пункте нужно выполнить три задания:

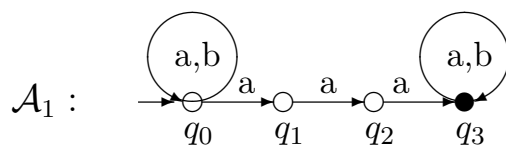
- 1) найти частное решение;
- 2) найти решение, минимальное по включению;
- 3) найти все решения.

1. $X = ((110)^* + 111^*)X.$

2. $X = (00 + 01 + 10 + 11)X + (0 + 1 + \varepsilon).$

3.
$$\begin{cases} Q_0 = 0Q_0 + 1Q_1 + \varepsilon, \\ Q_1 = 1Q_0 + 0Q_2, \\ Q_2 = 0Q_1 + 1Q_2. \end{cases}$$

Задача 11. Автомат \mathcal{A}_1 задан диаграммой. Выполните следующие задания. Достаточно выполнить хотя бы один из пунктов 2 или 3.



1. По диаграмме \mathcal{A}_1 постройте праволинейную грамматику G .
2. Запишите определяющую систему уравнений для G . Найдите её наименьшую неподвижную точку и вычислите регулярное выражение α_1 для $L(\mathcal{A}_1)$.

3. Определите регулярное выражение α_2 для $L(\mathcal{A}_1)$ с помощью индуктивного вычисления множеств R_{ij}^k .
4. Выберите какое-нибудь регулярное выражение α_1 или α_2 и постройте НКА \mathcal{A}_2 по регулярному выражению.
5. Выберите какой-нибудь НКА \mathcal{A}_1 или \mathcal{A}_2 и постройте ДКА D_1 .
6. Выберите какое-нибудь регулярное выражение α_1 или α_2 и постройте ДКА D_2 .
7. Выберите какой-нибудь ДКА D_1 или D_2 , дополните его, если нужно, до полного и постройте минимальный полный ДКА $\min \mathcal{A}$ для L . Для каждой пары состояний укажите соответствующие различающие их цепочки.
- 8*. По алгоритму КМП (Кнута–Мориса–Пратта) постройте автомат для L и сравните его с $\min \mathcal{A}$.

Контрольные вопросы

Несмотря на название раздела, все решения задач должны быть строго обоснованы.

Задача 12. Верно ли, что если пересечение языков $L_1, L_2 \subseteq \{a, b\}^*$ содержит язык $F = \{a^n b^n \mid n \geq 1\}$: $F \subseteq L_1 \cap L_2$, то хотя бы один из языков L_1 и L_2 является нерегулярным?

Задача 13. Пусть $X_1, X_2, \dots, X_n, \dots$ бесконечное семейство регулярных языков.

1. Верно ли, что язык $X = \bigcup_{n=1}^{\infty} X_n$ является регулярным языком?
2. Верно ли, что язык $X = \bigcap_{n=1}^{\infty} X_n$ является регулярным языком?

Задача 14. К языку L_1 добавили конечный язык R и получили язык L ($L = L_1 \cup R$). Язык L оказался регулярным. Верно ли, что язык L_1 мог быть нерегулярным?

Задача 15. Язык задан контекстно-зависимой грамматикой, которая не является контекстно-свободной. Может ли он быть регулярным?

Контекстно-свободные языки

Задача 16. Язык $L^=$ является языком всех слов с равным числом символов a и b .

1. Покажите индукцией³ по длине слова, что КС-грамматика с правилами $S \rightarrow SS \mid aSb \mid bSa \mid \varepsilon$ порождает язык $L^=$.
2. Грамматика называется линейной, если в правые части правил вывода входит не более одного нетерминала. Покажите, что язык $L^=$ не порождается никакой линейной КСГ.

Задача 17. Палиндромами называют слова, которые одинаково читаются слева направо и справа налево, например, «ротор».

1. Покажите, что язык палиндромов в произвольном алфавите является КС-языком.
2. Покажите, что дополнительный язык (язык всех непалиндромов) также является КС-языком.
3. Покажите, что дополнительный язык к языку $U = \{a^n b^n c^n, n = 0, 1, \dots\}$ является КС-языком.⁴

Задача 18. Являются ли следующие языки КС-языками?

1. $\{\{a, b\}^* \setminus ww \mid w \in \{a, b\}^*\}$.
2. $\{a^{3^n} \mid n > 0\}$.

³Другие доказательства, кроме индукции, не принимаются.

⁴Так как сам язык U не является КСЯ, то это означает, что в отличие от регулярных языков множество КСЯ не замкнуто относительно дополнения.

Задача 19. Выполните следующие задания.

1. Постройте магазинный автомат (МА), принимающий язык L^- из задачи 16.

2*. Постройте детерминированный МА, принимающий тот же язык, и приведите доказательство его корректности по индукции.

Задача 20. Язык Дика с двумя типами скобок D_2 порождается грамматикой

$$S \rightarrow SS \mid (S) \mid [S] \mid \varepsilon.$$

1. Постройте недетерминированный МП-автомат, распознающий язык D_2 .

2. Постройте детерминированный МП-автомат, распознающий язык D_2 , и приведите доказательство его корректности по индукции.

Задача 21. Для языка

$$L = \{w \mid w = xsy; x, y \in \{a, b\}^*; |x| = |y|\}$$

1) постройте КС-грамматику G , порождающую язык L ;

2) постройте недетерминированный МА, эквивалентный этой грамматике;

3) продемонстрируйте работу построенного МА на слове $acab$ (проанализируйте все варианты поведения).

Задача 22. Заданы грамматика $G = \{ \{A, B, C, D, E, F, S\}, \{a, b\}, \{S \rightarrow AB \mid C, A \rightarrow aE \mid a, E \rightarrow aE \mid \varepsilon, B \rightarrow bB \mid Bb \mid b, C \rightarrow CD, F \rightarrow ab, D \rightarrow aba\}, S \}$ и магазинный автомат $M = (\{q_0\}, \{a, b\}, \{S, a, b, A, B\}, \{ \delta(q_0, \varepsilon, S) = \{(q_0, AB)\}, \delta(q_0, \varepsilon, A) = \{(q_0, aA), (q_0, a)\}, \delta(q_0, \varepsilon, B) = \{(q_0, bB), (q_0, b)\}, \delta(q_0, a, a) = \{(q_0, \varepsilon)\}, \delta(q_0, b, b) = \{(q_0, \varepsilon)\}, q_0, S \}$, принимающий слова опустошением магазина.

1. Эквивалентны ли грамматика G и N -автомат⁵ M ?

⁵Мы называем N -автоматом МП-автомат, допускающий по пустому стеку, а F -автоматом — МП-автомат, допускающий по принимающему состоянию.

2. Однозначна ли грамматика G ? Если нет, то постройте эквивалентную ей однозначную грамматику.

3. Является ли автомат M детерминированным? Если нет, постройте эквивалентный ему детерминированный МА.

Задача 23. Определим языки $L_1 = \Sigma^*aab\Sigma^*$, $\Sigma = \{a, b\}$, и $L = \{w \mid w \in \overline{L_1}, |w|_a \geq |w|_b\}$.

1. Является ли дополнение языка L КС-языком?
2. Является ли дополнение языка L регулярным языком?

Задача 24. Язык L задан КСГ: $S \rightarrow aSa \mid aSb \mid bSa \mid bSb \mid a$.

1. Является ли L регулярным языком?
2. Является ли дополнение L регулярным языком?
3. Является ли L КС-языком?
4. Является ли дополнение L КС-языком?

Задача 25. Язык L задан грамматикой G :

$$S \rightarrow aSb \mid A \mid B \mid \varepsilon, \quad A \rightarrow aAa \mid \varepsilon, \quad B \rightarrow bBb \mid \varepsilon.$$

1. Является ли L регулярным языком?
2. Является ли дополнение L регулярным языком?
3. Является ли L КС-языком?
4. Является ли дополнение L КС-языком?

Контрольные вопросы

Задача 26. КС-грамматика называется *левооднозначной*, если каждое слово порождаемого ею языка имеет единственный левый вывод. Аналогично определяется *правооднозначная грамматика*. Можно ли построить пример левооднозначной, но не правооднозначной КС-грамматики?

Задача 27. Пусть L_1 – КС язык, не являющийся регулярным, а L_2 – не КС-язык. Может ли язык L_2L_1 быть регулярным языком? При положительном ответе привести пример.

Элементы синтаксического анализа

LL-анализ

Задача 28. Определить, являются ли $LL(k)$ -грамматиками следующие грамматики. Если да, указать точное значение k :

- а) $S \rightarrow Ab$, $A \rightarrow Aa \mid a$;
б) $S \rightarrow Ab$, $A \rightarrow aA \mid a$;
в) $S \rightarrow aAb$, $A \rightarrow BB$, $B \rightarrow ab \mid A \mid \varepsilon$;
г) $S \rightarrow aAb$, $A \rightarrow AaAb \mid \varepsilon$;
д) $S \rightarrow aB$, $B \rightarrow aBB \mid b$.

Задача 29. Построить $LL(1)$ -грамматику, эквивалентную грамматике из задачи **26(б)**, и управляющую таблицу для неё.

Задача 30. Написать для грамматики эквивалентную $LL(1)$ -грамматику, построить $LL(1)$ -анализатор и продемонстрировать его работу на слове $baab$.

$$S \rightarrow baaA \mid babA \quad A \rightarrow \varepsilon \mid Aa \mid Ab$$

Задача 31*. Докажите, что язык $a^* \cup a^n b^n$ не является LL(1)-языком, то есть не существует LL(1)-грамматики, порождающей этот язык.

Задача 32. Язык L задан неоднозначной КС-грамматикой

$$G = \{\{S\}, \{(,)\}, \{S \rightarrow (S) \mid SS \mid ()\}, S\}.$$

Написать LL(1)-грамматику для языка L .

Контрольные вопросы

Задача 33. Существует ли такая праволинейная (не обязательно регулярная праволинейная) грамматика, которая не является LL(1)-грамматикой?

Задача 34. В приведённой грамматике⁶ G есть правило $S \rightarrow AB$ и при этом $\text{FIRST}(A) \cap \text{FIRST}(B) = \varepsilon$. Верно ли, что грамматика G может быть LL(1)-грамматикой?

Задача 35. Пусть для некоторых двух нетерминалов A и B приведённой КС-грамматики G пересечение $\text{FOLLOW}(A) \cap \text{FOLLOW}(B)$ оказалось непустым. Верно ли, что грамматика G не является LL(1)-грамматикой?

⁶Грамматика называется приведённой, если в ней нет недостижимых и бесплодных символов. В литературе также встречаются неэквивалентные определения этого термина.

LR-анализ

Задача 36. Дана грамматика $G = \{ \{A, S\}, \{a, b, c\}, \{ S \rightarrow Aa \mid b \mid \varepsilon; A \rightarrow Ab \mid c \}, S \}$. Является ли грамматика G LR(k)-грамматикой? При положительном ответе на вопрос найти минимальное k и построить соответствующий анализатор. Построить дерево разбора для цепочки $cbba$.

Задача 37. Дана грамматика $G = \{ \{A, S\}, \{a\}, \{ S \rightarrow A; A \rightarrow aAa \mid a \}, S \}$. Является ли грамматика G LR(k)-грамматикой? При положительном ответе на вопрос найти минимальное k и построить соответствующий анализатор. Построить дерево разбора для цепочки $aaaaa$.

Задача 38. Дана грамматика $G = \{ \{A, S\}, \{a, b, c\}, \{ S \rightarrow Aa \mid b; A \rightarrow Ab \mid c \}, S \}$. Является ли грамматика G LR(k)-грамматикой? При положительном ответе на вопрос найти минимальное k и построить соответствующий анализатор. Продемонстрировать работу анализатора на цепочке $cbbab$.

Задача 39. Зафиксируем КС-грамматику G и рассмотрим множество её LR(0)-ситуаций. Будем говорить, что между двумя ситуациями $\alpha.X\beta$ и $\alpha X.\beta$ определён переход по $X \in N \cup T$. Конечный автомат, в качестве состояний которого выступают LR(0)-ситуации, а переходы определены по правилу, указанному выше, называют LR(0)-автоматом или автоматом Кнута.

1. Выпишите все LR(0)-ситуации для грамматики G , заданной правилами $S \rightarrow aS \mid b$.
2. Постройте автомат Кнута для грамматики G .
3. Постройте LR(0)-анализатор для грамматики G . Сравните автомат Кнута с таблицей переходов LR(0)-анализатора для грамматики G .

Задача 40. Грамматика G задана правилами:

$$S \rightarrow Ab, \quad A \rightarrow aAa, \quad A \rightarrow B, \quad B \rightarrow b.$$

1. Построить LR(1) и LR(0)-анализаторы для грамматики G по алгоритму из курса.
2. Постройте LR(0)-анализатор по LR(1)-анализатору из пункта 1 следующим образом. Сотрите все аванцепочки и постройте управляющую таблицу LR(0)-анализатора по получившемуся автомату Кнута. Верно ли, что полученный LR(0)-анализатор является анализатором для грамматики G ? То есть для любого слова, порождаемого G , анализатор строит корректный правый разбор, а слова, не порождаемые G , анализатор отвергает.
3. Покажите, что LR(0)-анализатор для грамматики G из пункта 1 можно построить путём применения следующей процедуры, схожей с процедурой минимизации ДКА, к LR(0)-автомату, полученному из LR(1)-анализатора в пункте 2.

В случае минимизации LR(0)-автомата, все состояния с операциями свёртки оказываются на первом шаге в разных группах (разных ящиках), если свёртки происходят по разным правилам; состояния с операциями сдвига находятся в одном ящике. Далее процедура минимизации LR(0)-автомата не отличается от процедуры минимизации ДКА.

Контрольные вопросы

Задача 41. При построении LR(1)-анализатора для грамматики G в одном множестве оказались ситуации $[A \rightarrow .aA\alpha, b]$ и $[B \rightarrow \beta.a, a]$, где α, β некоторые цепочки из $(N \cup T)^*$. Может ли грамматика G оказаться LR(0)-грамматикой?

Атрибутные грамматики

Атрибутные грамматики являются исключительно важными для понимания роли всего изученного материала (теории) в процессе реализации компиляторов для языков программирования. Однако их изучение приходится на конец курса, поэтому мы приводим в этом разделе теоретический материал, подготовленный С.П. Тарасовым, для облегчения

изучения небольшой части этой значительной темы, которая входит в наш курс.

Об атрибутивных грамматиках можно прочитать в книге В.А. Серебрякова. Они встречались в дополнительных вопросах на экзамене. Неформально, определение атрибутов для данной КС-грамматики G приписывает каждому выводу в G некоторое *индуктивное вычисление*. Поскольку индуктивное вычисление по выводу (дереву вывода) может идти как снизу вверх, так и сверху вниз, то технически правилам вывода G приписываются т.н. *синтезируемые атрибуты, вычисляемые снизу вверх через атрибуты потомков* и *наследуемые атрибуты, вычисляемые сверху вниз через атрибуты предков*.

Говоря формально⁷, с каждым символом $X \in V$ связывается конечное множество атрибутов $A(X)$, которое разбивается на два непересекающихся множества: множество синтезированных атрибутов $A_0(X)$ и множество унаследованных атрибутов $A_1(X)$. Множество $A_1(S)$ должно быть пустым (то есть начальный символ S не должен иметь унаследованных атрибутов). Аналогично, множество $A_0(X)$ пусто, если X — терминальный символ. Каждый атрибут α из множества $A(X)$ имеет (возможно, бесконечное) множество значений $V\alpha$. Для каждого вхождения X в дерево вывода семантические правила позволяют определить одно значение из множества $V\alpha$ для соответствующего атрибута.

Пусть G состоит из m правил, и пусть p -е правило имеет вид $X_{p0} \rightarrow X_{p1}X_{p2}\dots X_{pn_p}$, где $n_p \geq 0$, $X_{p0} \in N$ и $X_{pj} \in V$ для $1 \leq j \leq n_p$. Семантическими правилами называются функции $F_{pj\alpha}$, определённые для всех $1 \leq p \leq m$, $0 \leq j \leq n_p$ и некоторых $\alpha \in A_0(X_{pj})$, если $j = 0$ или $\alpha \in A_1(X_{pj})$, если $j > 0$. Каждая такая функция представляет собой отображение из $V\alpha_1 \times V\alpha_2 \times \dots \times V\alpha_t$ в $V\alpha$ для некоторого $t = t(p, j, \alpha) \geq 0$, где все $\alpha_i = \alpha_i(p, j, \alpha)$ являются атрибутами некоторых X_{pk_i} при $0 \leq k_i = k_i(p, j, \alpha) \leq n_p$, $1 \leq i \leq t$. Другими словами, каждое семантическое правило отображает значения некоторых атрибутов символов $X_{p0}, X_{p1}, \dots, X_{pn}$ и значение некоторого атрибута символа X_{pj} .

С одной стороны, задание атрибутов удобно для моделирования семантики языков программирования. Однако, это вычислительное средство является настолько мощным, что уже простейшие проверки корректности системы атрибутов практически трудно реализуемы (см. при-

⁷Далее идет заимствование из оригинальной статьи Д. Кнута, перевод которой приведён в книге В.А. Серебрякова

ложения **A** и **B** в книге В.А. Серебрякова). Это в значительной степени ограничивает их применение. Тем не менее, мне кажется, будет очень полезно ознакомиться со статьей Д. Кнута (приложение **A** в книге Серебрякова) и иметь в виду, что *незацикленность* является разрешимой, хотя и очень трудоемкой (экспоненциальной по входу) задачей.

Рассмотрим грамматику⁸

$$G = \{\{S, L, B\}, \{0, 1\}, \{S \rightarrow L \mid L.L, L \rightarrow B \mid LB, B \rightarrow 0 \mid 1\}, S\}.$$

В грамматике G можно вывести произвольные двоичные числа (нетерминалы B (bit) и L (list) интерпретируются, соответственно, как *бит* и *последовательность битов*). Рассмотрим два варианта *атрибутов*, позволяющих в процессе вывода вычислять десятичное значение выводимого числа.

Список атрибутов 1

- B имеет целочисленный атрибут “значение”, обозначаемый $v(B)$.
- L имеет целочисленные атрибуты “длина”, обозначаемый $l(B)$, и “значение”, обозначаемый $v(L)$.
- S имеет атрибут “значение”, являющийся рациональным числом и обозначаемый $v(N)$.

Семантические правила 1

$$\begin{array}{lll} B \rightarrow 0 & v(B) = 0 & \\ B \rightarrow 1 & v(B) = 1 & \\ L \rightarrow B & v(L) = v(B), & l(L) = 1 \\ L \rightarrow LB & v(L_1) = 2v(L_2) + v(B), & l(L_1) = l(L_2) + 1 \\ S \rightarrow L & v(S) = v(L) & \\ S \rightarrow L.L & v(S) = v(L_1) + v(L_2)/2^{l(L_2)} & \end{array}$$

(Индексы в четвёртом и шестом правилах применяются для того, чтобы различать вхождения одноимённых нетерминалов.)

⁸Взята из уже цитированной статьи Д. Кнута, помещенной в виде приложения к книге В.А. Серебрякова.

Список атрибутов 2

- B имеет рациональный атрибут “значение”, обозначаемый $v(B)$, и целочисленный атрибут “масштаб”, обозначаемый $s(B)$.
- L имеет рациональный атрибут “значение”, обозначаемый $v(L)$, целочисленный атрибут “длина”, обозначаемый $l(L)$, и целочисленный атрибут “масштаб”, обозначаемый $s(L)$.
- N имеет рациональный атрибут “значение”, обозначаемый $v(N)$.

Семантические правила 2.

$$\begin{array}{ll} B \rightarrow 0 & v(B) = 0 \\ B \rightarrow 1 & v(B) = 2^{s(B)} \\ L \rightarrow B & v(L) = v(B), \quad s(B) = s(L), \\ & l(L) = 1 \\ L_1 \rightarrow L_2 B & v(L_1) = v(L_2) + v(B), \quad s(B) = s(L_1), \\ & s(L_2) = s(L_1) + 1, \quad l(L_1) = l(L_2) + 1 \\ S \rightarrow L & v(S) = v(L), \quad s(L) = 0 \\ S \rightarrow L_1.L_2 & v(S) = v(L_1) + v(L_2), \quad s(L_1) = 0, \\ & s(L_2) = -l(L_2) \end{array}$$

Здесь при записи семантических правил принято следующее соглашение. Правая часть каждого правила представляет собой определение левой части. Таким образом, $s(B) = s(L)$ означает, что сначала должно быть вычислено $s(L)$, а затем полученное значение следует присвоить $s(B)$.

Задача 42.

1. Перечислите синтезируемые и наследуемые атрибуты для обеих систем семантических правил.
- 2–3. Для каждой из описанных выше систем атрибутов и семантических правил вычислите десятичное значение двоичного числа 100.001001.

Дополнительные задачи

В этот раздел входят задачи для подготовки к контрольным работам и экзаменам, а также задачи повышенной сложности для студентов, претендующих на высокие оценки. Задачи данного раздела не являются обязательными для прохождения процедуры сдачи задания, если только не входят в требования семинариста. Во всех письменных общекурсовых работах значение k в задачах на построение LR(k)-анализаторов не превосходит единицу.

Регулярные языки

Задача 43. Пусть X регулярный язык. Верно ли, что язык $\bigcap_{n=1}^{\infty} (\Sigma^* \setminus X)^n$ является регулярным?

Задача 44. Приведите пример бесконечного регулярного языка $X \subset \{a, b\}^*$, отличного от множества всех слов, такого что $X \cap (\Sigma^* \setminus X)^R = X$.

Задача 45. Найдите разбиения на минимальное число классов правоинвариантной (И/ИЛИ левоинвариантной) эквивалентности, которые индуцируют следующие языки.

1. Язык, порождаемый выражением $00(10 + 01)^*$.
2. Язык $\{a^{n^2} \mid n \geq 0\}$ в однобуквенном алфавите.

КС-языки

Задача 46. Язык L задан грамматикой G :

$$S \rightarrow bSa \mid AB \mid \varepsilon, \quad A \rightarrow bAb \mid b, \quad B \rightarrow aBa \mid \varepsilon.$$

Является ли язык L и его дополнение регулярным языком, КС-языком?

Задача 47. Являются ли следующие языки КС-языками?

1. $\{x \mid x \in \{c, b\}^*, |x|_c = |x|_b, \forall u, v : x = uv, |u| \neq 0, |v| \neq 0, |u|_c > |u|_b\}$.
2. $\{a^{3^n} \mid n > 0\}$.

Задача 48*. Пусть – МА. Постройте МА B , принимающий все префиксы языка $L(A)$, т.е. язык $L(B) = \{x \mid \exists y : xy \in L(A)\}$.

Задача 49. Для языка

$$L = \{w \mid w = xc^{3k}y; x, y \in \{a, b\}^*; |xy|_a = 2n; n, k \geq 0\}$$

($|xy|_a$ – число символов a в слове xy)

- 1) постройте КС-грамматику G , порождающую язык L ;
- 2) постройте недетерминированный МА, эквивалентный этой грамматике;
- 3) продемонстрируйте работу построенного МА на слове $accab$ (проанализируйте все варианты поведения).

Задача 50. Заданы языки $L_1 = \{a^n b^n c^m : n \geq 1, m \geq 0\}$, $L_2 = \{f^n a^m b^m : n \geq 0, m \geq 0\}$. Для языка $L_1 \cup L_2$ построить однозначную КС-грамматику и детерминированный МП-автомат. Решение обосновать.

Элементы синтаксического анализа

Задача 51. Язык L задан неоднозначной КС-грамматикой:

$$G = \{ \{S\}, \{a, \cdot, \wedge, [,], (,)\}, \{S \rightarrow a \mid S.S \mid S[S] \mid S^\wedge \mid S(S)\}, S \}.$$

Написать LL(1)-грамматику для языка L .

Задача 52. Дана грамматика $G = \{ \{A, B, C, D, E, S\}, \{a, b\}, \{S \rightarrow AB, A \rightarrow a, B \rightarrow CD \mid aE, C \rightarrow ab, D \rightarrow bb, E \rightarrow bba\}, S \}$. Является ли грамматика G LR(k)-грамматикой? При положительном ответе на вопрос найти минимальное k и построить соответствующий анализатор. Продемонстрировать работу анализатора на цепочке $aabbb$.

Задание составил

А.А. Рубцов, старший преподаватель

С методическими материалами по курсам кафедры МОУ можно ознакомиться на страницах:

<http://www.mou.mipt.ru>, <http://trpl7.ru>,
<http://lrk.umeta.ru>, <http://rubtsov.su>.