

УТВЕРЖДАЮ
Проректор по учебной работе
Ю.А. Самарский

8 июня 2013 г.

ПРОГРАММА

по курсу ТЕОРИЯ И РЕАЛИЗАЦИЯ ЯЗЫКОВ
ПРОГРАММИРОВАНИЯ

по направлению 010900

факультет ФУПМ

кафедра математических основ управления

курс II

семестр 3

лекции – 34 часов Экзамен – 3 семестр

семинары – 34 часов

лабораторные занятия – нет

самостоятельная работа – 2 часа в неделю

ВСЕГО ЧАСОВ – 68

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

Программа обсуждена на заседании кафедры
математических основ управления

17 мая 2013 года

Заведующий кафедрой

ПРОГРАММА

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 автоматными грамматиками (грамматики типа 3: левосторонние – ЛЛ, правосторонние – 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.
- Алгоритм Ахо-Корасик.
- Построение по НКА эквивалентных ЛЛ- и ПЛ-грамматик.
- Построение эквивалентного НКА по ЛЛ- и ПЛ-грамматике.
- Решение уравнений с регулярными коэффициентами.

- Синтаксический анализ: КС-грамматики (КСГ). Преобразования КС-грамматик, приведённые грамматики. Канонические формы КС-грамматик (форма Хомского и форма Грейбах). Свойства замкнутости КС-языков (КСЯ), незамкнутость КСЯ относительно пересечения. Дерево вывода КСГ. Однозначность КС-грамматик. Однозначность правой линейной грамматики, построенной по ДКА. Лемма о накачке для КСЯ. Проверка принадлежности слова КСЯ КСГ (алгоритм Кока- Янгера- Касами)
3. Синтаксический анализ: автоматы с магазинной памятью (МА). Детерминированные и недетерминированные МА. Обобщённые МА, 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).
4. **Дополнительные сведения из теории конечных автоматов**
- 1.1. Минимизация числа состояний и эквивалентность детерминированного конечного автомата (ДКА).

5. Синтаксический анализ

- 2.1. Эквивалентность МА, распознающих по конечному состоянию и по опустошению магазина.
- 2.2. Преобразования КС-грамматик, приведённые грамматики.
- 2.3. Предсказывающий разбор *сверху вниз*.
- 2.3.1. Алгоритм разбора *сверху вниз*.
- 2.3.2. Функции *FIRST* и *FOLLOW*.
- 2.3.3. Конструирование таблицы предсказывающего анализатора.
- 2.3.4. LL(1)-грамматики.
- 2.3.5. Удаление левой рекурсии.
- 2.3.6. Левая факторизация.
- 2.3.7. Рекурсивный спуск.
- 2.3.8. LL(k)-грамматики.
- 2.4. Разбор *снизу вверх* типа сдвиг-свёртка.
- 2.4.1. Основа.
- 2.4.2. LR(1)-анализаторы.
- 2.4.3. Конструирование LR(1)-таблицы.
- 2.4.4. LR(1)-грамматики.
- 2.4.5. Варианты LR-анализаторов.
- 2.4.6. LR(k)-грамматики.

6. Элементы теории перевода

- 3.1. Синтаксически управляемый перевод.
- 3.2. Атрибутные грамматики.

Литература

1. Ахо А., Сети Р., Ульман Дж. Компиляторы. Принципы, технологии, инструменты. – М.; СПб.; Киев; Вильямс, 2001 (681.3 / А95).
2. Мартыненко Б.К. Языки и трансляции. – СПб.: СПбГУ, 2004. <http://www.math.spbu.ru/user/mbk/TUTORY/LT.html>
3. Серебряков В.А. [и др.]. Теория и реализация языков программирования. – М.: МЗ-Пресс, 2006. – 294 с. (681.3/ Т33).

4. Хопкрофт Дж., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. – М.: Вильямс, 2002.(681.3 / X78).
5. Ахо А., Лам М., Сети Р., Ульман Дж. Компиляторы. Принципы, технологии и инструментарий. М.; СПб.; Киев: Вильямс, 2011. – 1184 с. (681.3/ А95).

ЗАДАНИЕ

Метасимволы * и ? обозначают соответственно любое слово и любой символ в данном алфавите.

Обращение слова (или языка) u обозначается через u^R .

Длина слова x обозначается через $|x|$.

Число букв a в слове x обозначается через $|x|_a$.

Запись w^n обозначает слово $\underbrace{ww\dots w}_n$.

«Построить объект» означает, что нужно предъявить сам объект и доказать, что объект удовлетворяет всем условиям задачи.

Конечные автоматы и регулярные языки

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

- (i) $\varepsilon, b, bb \in L$;
- (ii) вместе с любым словом $x \in L$ в L также входят слова $ax, bax, bbax$;
- (iii) никаких других слов в L нет.

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

Докажите или опровергните, что $L = T$.

Если равенство неверно, то нужно явно указать слово, принадлежащее одному языку и не принадлежащее другому.

Если равенство верно, то нужно провести доказательство по индукции:

$$L \subseteq T; (L \supseteq T \text{ (т.е. } L = T)).$$

2.1. Верно ли, что

(i) $\varepsilon \in \{a, aab, aba\}$?

(ii) $\emptyset \in \{a, aab, aba\}$?

2.2. Постройте РЯ:

(iii) $\{a^n \mid n > 0\} \times \{b^n \mid n \geq 0\} = ?$

(iv) $\{a^{3n} \mid n > 0\} \cap \{a^{5n+1} \mid n \geq 0\}^* = ?$

3. Пусть L – произвольный язык в алфавите $\{a, b\}$. Пусть язык

$$L_1 = \{xca^m b^n \mid x \in L, m \leq 2n, m \geq 0, n > 1\}$$
 в алфавите $\{a, b, c\}$.

Явно постройте дополнение к языку L_1 , т.е. представьте дополнение в виде объединения известных языков.

В качестве операнда можно использовать язык $\{a, b\}^* \setminus L$.

4. КА A задан таблицей (конечное состояние – q_1):

	q_0	q_1	q_2	q_3
0	q_3	q_2	q_1	q_0
1	q_1	q_0, q_3	-	-

i. Является ли A детерминированным?

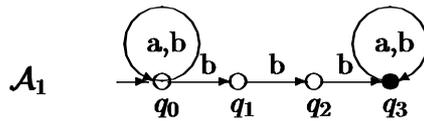
ii. Опишите все последовательности конфигураций A при обработке слова $w=0011001$. Верно ли, что $w \in L(A)$?

iii. Укажите два слова, принадлежащие языку $L(A)$, и два слова, не принадлежащие языку $L(A)$

5. i. Построить ДКА, принимающий язык L всех слов в алфавите $\{0,1\}$, содержащих чётное число нулей и нечётное число единиц.

ii. Построить эквивалентную левостороннюю грамматику. Будет ли она однозначной?

- iii. Построить регулярное выражение для языка L^R .
6. Будут ли регулярными следующие языки
- i. $L_1 = \{a^{2013n+5}, n = 0, 1, \dots\} \cap \{a^{503k+29}, k = 401, 402, \dots\} \subseteq \{a^*\}$
- ii. $L_2 = \{a^{200n^2+1}, n = 1000, 1001, \dots\} \subseteq \{a^*\}$
- iii. 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, $001010 (1010_2 = 10_{10} = 3 \times 3 + 1) \notin L_3$,
 $a 10001 (10001_2 = 17_{10} = 5 \times 3 + 2) \in L_3$.
7. Постройте НКА, принимающий язык $L_k = \{\text{Множество слов в алфавите } \{0, 1\}, \text{ у которых третий от конца символ равен } \langle 1 \rangle\}$.
 Затем, используя алгоритм, постройте соотв. полный ДКА.
8. Решите уравнения с регулярными коэффициентами (найти частное решение; найти все решения; найти минимальное по включению решение):
- i. $X = ((101)^* + 110^*) X$.
- ii.
$$\begin{cases} X_1 = (1 + (01)^*) X_1 + (0 + 11)^* X_2 + X_3, \\ X_2 = 0X_1 + X_2 + (110)^* X_3, \\ X_3 = 1X_1 + 10X_2 + 1. \end{cases}$$
9. Пусть язык $L \subseteq \{a, b\}^*$ задан следующей диаграммой.



- i. По диаграмме \mathcal{A}_1 постройте праволинейную грамматику G для L .
 - ii. Запишите определяющую систему уравнений для G . Найдите её наименьшую неподвижную точку и вычислите регулярное выражение α_1 для L .
 - iii. Определите регулярное выражение α_2 для L с помощью индуктивного вычисления множеств R_{ij}^k (см. § 3.4 книги [3]).
 - iv. Выберите какое-нибудь из регулярных выражений α_1 или α_2 и постройте НКА \mathcal{A}_2 по регулярному выражению (см. § 3.3.1 книги [3]).
 - v. Выберите какой-нибудь из НКА \mathcal{A}_1 или \mathcal{A}_2 и постройте ДКА D_1 .
 - vi. Выберите какое-нибудь из регулярных выражений α_1 или α_2 и постройте ДКА D_2 (см. § 3.3.3 книги [3]).
 - vii. Выберите какой-нибудь из ДКА D_1 или D_2 , дополните его, если нужно, до полного и постройте минимальный полный ДКА $\min A$ для L . Для каждой пары состояний укажите соответствующие различающие цепочки.
 - viii. Постройте КМП-автомат (КМП значит Кнута-Мориса-Пратта) для L и сравните его с $\min A$.
10. Порождает ли выражение $(ab)^*(ba)^*$ тот же язык, что распознаётся ДКА $M = (\{A,B,C,D\}, \{a,b\}, \Delta, A, \{A,D,E\})$, где функция переходов задана следующим образом:

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

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

Следующая задача необязательна

11. Найдите разбиение на минимальное число классов правоинвариантной (И/ИЛИ левоинвариантной) эквивалентности, которые индуцируют следующие языки.

- i. Язык, порождаемый выражением $00(10+01)^*$.
- ii. Язык $\{a^{n^2}\}$ в однобуквенном алфавите.

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

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

13. i. Покажите индукцией по длине слова, что КСГ с правилами $S \rightarrow SS \mid aSb \mid bSa \mid \epsilon$ порождает язык $L^=$ всех слов с равным числом литералов a и b .

Другие доказательства, кроме индукции, не принимаются.

ii. Грамматика называется *линейной*, если в правые части правил вывода входит не более одного нетерминала. Покажите, что язык $L^=$ не порождается никакой линейной КСГ.

14. i. Покажите, что язык палиндромов в произвольном алфавите (т.е. язык всех слов, которые одинаково читаются слева направо и справа налево, например, «ротор») является КС-языком.

ii. Покажите, что дополнительный язык всех непалиндромов также является КС-языком.

iii. Покажите, что дополнительный язык к языку $U = \{a^n b^n c^n, n = 0, 1, \dots\}$ является КС-языком. Так как сам язык U не является КСЯ, то это означает, что в отличие от регулярных языков множество КСЯ не замкнуто относительно дополнения.

15. 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?

16. Постройте однозначную КС-грамматику (однозначность нужно доказать) языка $\{x \mid x \in \{c, b\}^*, |x|_c = |x|_b, \text{ и для } \forall u, v: x = uv, |u| \neq 0, |v| \neq 0, |u|_c > |u|_b\}$.

17. Являются ли следующие языки КС-языками:

i. $\{\{a, b\}^* \setminus ww \mid w \in \{a, b\}^*\}$;

ii. $\{a^{3^n} \mid n > 0\}$

18. **i.** Постройте магазинный автомат (МА), принимающий язык L^- из задачи 13.

ii. Постройте детерминированный МА, принимающий тот же язык и приведите доказательство его корректности по индукции (моделирование счётчика).

19. Пусть A – МА. Постройте МА B , принимающий все префиксы языка $L(A)$, т.е. язык

$$L(B) = \{x \mid xy \in L(A)\}.$$

20. Для языка

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

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

i. постройте КС-грамматику;

ii. постройте недетерминированный МА, эквивалентный этой грамматике;

iii. продемонстрируйте работу построенного МА на слове *accab* (проанализируйте все варианты поведения).

21. 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 МА $M = (\{q_0\}, \{a, b\}, \{S, a, b, A, B\}, \{\delta(q_0, \varepsilon, S) = \{(q_0, AB)\}, \delta(q_0, \varepsilon, A) = \{(q_0, aA), (q_0, a)\}, \delta(q_0, \varepsilon, B) = \{(q_0, bB), (q_0, b)\}, \delta(q_0, a, a) = \{(q_0, \varepsilon)\}, \delta(q_0, b, b) = \{(q_0, \varepsilon)\}, q_0, S\}$,
принимаящий слова опустошением магазина.

i. Эквивалентны ли грамматика G и N -автомат M ?

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

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

22. 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.

23. $L_1 = (a \mid b)^* aab(a \mid b)^*$. Язык $L = \{w \mid w \in \overline{L_1}, |w|_a \geq |w|_b\}$.

1. Является ли дополнение языка L КС-языком?

2. Является ли дополнение L регулярным языком?

Часть 2.

LL-грамматики

1. Определить, являются ли следующие грамматики $LL(k)$ -грамматиками. Если да, указать точное значение k :
 - а) $S \rightarrow Ab, A \rightarrow Aa | a$;
 - б) $S \rightarrow Ab, A \rightarrow aA | a$;
 - в) $S \rightarrow aAb, A \rightarrow BB, B \rightarrow ab | A | \epsilon$;
 - г) $S \rightarrow aAb, A \rightarrow AaAb | \epsilon$;
 - д) $S \rightarrow aB, B \rightarrow aBB | b$.
2. Построить $LL(1)$ -грамматику, эквивалентную грамматике из задачи (3б), и управляющую таблицу для неё.
3. Для грамматики из задачи (3д) построить управляющую таблицу. Одинаковые ли языки порождают грамматики из (3в), (3г), (3д)?

LR-грамматики

4. Являются ли следующие грамматики $LR(k)$ -грамматиками (указать точное значение k и построить соответствующий детерминированный правый анализатор):
 - а) грамматики, порождающие языки из задач (5.i и 7 части I), построенные вами ранее (подсказка: посмотрите, нельзя ли упростить вашу грамматику);
 - б) грамматики из задачи (3).

Атрибутные грамматики

5. В грамматике $[целое] \rightarrow aC, C \rightarrow aC | \epsilon$ терминал a имеет атрибут 0 или 1. Определить атрибуты так, чтобы нетерминал $[целое]$ имел атрибут, равный восьмеричному значению выводимого числа.
6. Построить атрибутные грамматики для следующих переводов:

$$\{(x, x) | x \in \{a, b\}^*\};$$

$$\left\{ (x, x^R) \mid x \in \{a, b\}^* \right\};$$

$$\left\{ (x, xx) \mid x \in \{a, b\}^* \right\};$$

$$\left\{ (a^n b^n, c^n b^n a^n) \mid n \geq 1 \right\}.$$

7. Дополнить грамматику $S \rightarrow 1S00, S \rightarrow 0S11, S \rightarrow \varepsilon$ до атрибутной так, чтобы вычислялась максимальная длина непрерывной последовательности единиц в порождённом слове.
8. Дополнить грамматику $S \rightarrow AA, A \rightarrow A0, A \rightarrow A1, A \rightarrow \varepsilon$ до атрибутной так, чтобы вычислялось число сочетаний «01» в порождённом слове.

С методическими материалами по курсам кафедры МОУ вы можете познакомиться на страницах:
www.mou.mipt.ru, <http://trpl.narod.ru>
<http://lrk.umeta.ru>

Подписано в печать 08.06.2012. Формат 60 × 84¹/₁₆. Бумага офсетная.
 Печать офсетная. Усл. печ. л. 0,44. Тираж 370 экз. Заказ № 33

Государственное образовательное учреждение
 высшего профессионального образования
 Московский физико-технический институт
 (государственный университет)
 Отдел оперативной полиграфии
 141700, Моск. обл., г. Долгопрудный, Институтский пер., 9.