

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Формальные языки и трансляции
по направлению:	Прикладная математика и информатика
профиль подготовки:	А1360: Передовые методы искусственного интеллекта Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 5 (осенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Количество контрольных работ, заданий: 1

Программу составил: А.А. Сорокин, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры алгоритмов и технологий программирования 04.06.2020

Аннотация

В курсе рассматриваются основные понятия теории формальных языков. Прежде всего, вводятся базовые понятия алфавита, слова и языка, после чего рассматриваются различные способы задания множеств слов (языков). Среди таких способов выделяются два основных: конечные автоматы и контекстно-свободные грамматики. Также рассматриваются эквивалентные перечисленным способы описания: регулярные выражения и праволинейные грамматики, задающие автоматные языки, и автоматы с магазинной памятью, эквивалентные контекстно-свободным грамматикам. Кроме того, в курсе подробно рассматриваются конечные преобразователи. Курс содержит большое количество математических теорем, описывающих свойства введенных объектов и соотношения между ними. Также в курсе подробно разбираются алгоритмы синтаксического анализа для контекстно-свободных грамматик, их алгоритмическая сложность и условия применения.

Курс содержит большое количество теоретического материала, задач, а также практические задания на реализацию алгоритмов. Для успешного освоения курса слушателю необходимо изучить дисциплины «Математическая логика и теория алгоритмов» и «Алгоритмы и структуры данных», а также владеть основными методами доказательства математических утверждений.

1. Цели и задачи

Цель дисциплины

знакомство студентов с основными разделами теории формальных языков для последующего использования в других дисциплинах естественнонаучного содержания и практической деятельности, развитие математической культуры, исследовательских и программистских навыков.

Задачи дисциплины

Заложить базовые знания в области теории формальных языков.

Развить общематематическую культуру: умение логически мыслить, формулировать и доказывать строгие математические утверждения.

Научить выбирать алгоритм для решения задачи, обосновывать его правильность и реализовывать на требуемом языке программирования.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- понятия формального языка и грамматики;
- классификацию формальных языков и грамматик;
- регулярные выражения;
- детерминированные и недетерминированные конечные автоматы;
- контекстно-свободные грамматики;
- основные алгоритмы синтаксического разбора для контекстно-свободных грамматик;
- автоматы с магазинной памятью;
- важнейшие подклассы контекстно-свободных грамматик (в т. ч. LR(k)-грамматики);
- конечные преобразователи
- практическое применение автоматов и грамматик, в т. ч. в задачах компиляции.

уметь:

делать правильные выводы из сопоставления результатов теории и эксперимента;
 определять класс грамматик, необходимый для решения задачи;
 формально доказывать соответствие формальных языков и грамматик поставленной задаче;
 строить конечный автомат/регулярное выражение по описанию автоматного языка;
 строить регулярное выражение по заданному конечному автомату;
 строить конечный автомат по заданному регулярному выражению;
 строить контекстно-свободную грамматику по описанию контекстно-свободного языка;
 строить автомат с магазинной памятью для заданной контекстно-свободной грамматики;
 строить синтаксический анализатор для заданной контекстно-свободной грамматики с помощью генераторов синтаксических анализаторов.

владеть:

математическим аппаратом теории формальных языков;
 навыками самостоятельной работы в современных программных комплексах;
 навыками освоения большого объема информации;
 навыками использования генераторов синтаксических анализаторов.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Задачи, цели и методы теории формальных языков.	2	2		5
2	Регулярные выражения.	2	2		5
3	Конечные автоматы.	6	6		7
4	Контекстно-свободные грамматики.	8	8		5
5	Алгоритмы синтаксического анализа.	6	6		5
6	Конечные преобразователи.	4	4		8
7	Приложения теории формальных языков.	2	2		10
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 5 (Осенний)

1. Задачи, цели и методы теории формальных языков.

Теория формальных языков как научная дисциплина. История теории формальных языков, её историческое значение для лингвистики и программирования. Философское значение теории формальных языков. Краткое описание современных задач.

2. Регулярные выражения.

Понятие регулярного выражения и регулярного языка. Замкнутость класса регулярных языков относительно основных операций над языками. Отдельные примеры регулярных языков.

3. Конечные автоматы.

Недетерминированные конечные автоматы. Различные варианты их определения. Эквивалентность этих вариантов. Эпсилон-переходы и их устранение. Детерминированные конечные автоматы. Совпадение классов языков, задаваемых детерминированными и недетерминированными автоматами. Алгоритм детерминизации конечного автомата. Теорема Клини о совпадении классов регулярных и автоматных языков. Алгоритм построения конечного автомата по заданному регулярному выражению. Понятие минимального автомата. Минимизация детерминированных конечных автоматов. Алгоритм проверки равенства регулярных выражений. Лемма о разрастании для автоматных языков. Примеры неавтоматных языков.

4. Контекстно-свободные грамматики.

Понятие порождающей грамматики. Иерархия Хомского. Праволинейные грамматики. Совпадение классов языков, задаваемых праволинейными грамматиками и конечным автоматами. Понятие контекстно-свободной грамматики и контекстно-свободного языка. Вывод в контекстно-свободной грамматике и деревья вывода. Понятие однозначной грамматики. Замкнутость класса контекстно-свободных языков относительно отдельных операций. Нормальные формы для контекстно-свободных языков. Устранение эпсилон-правил в контекстно-свободных грамматиках. Нормальные формы Хомского и Грейбах. Автоматы с магазинной памятью, варианты определения. Эквивалентность различных вариантов. Автомат с магазинной памятью в нормальной форме. Совпадение классов языков, задаваемых автоматами с магазинной памятью и контекстно-свободными грамматиками. Лемма о разрастании для контекстно-свободных языков. Примеры языков, не являющихся контекстно-свободными. Детерминированные автоматы с магазинной памятью, различные варианты определения. Детерминированные контекстно-свободные языки. Однозначность детерминированных контекстно-свободных языков. Свойства класса детерминированных контекстно-свободных языков. Линейные и полулинейные множества, теорема Парика.

5. Алгоритмы синтаксического анализа.

Алгоритм Кока-Янгера-Касами для контекстно-свободной грамматики в нормальной форме Хомского, алгоритм Эрли синтаксического разбора, алгоритм восходящего синтаксического анализа «перенос-свёртка», LR-таблица и алгоритм синтаксического разбора по LR-таблице, понятие LR-ситуации, состояния LR-анализатора, алгоритм построения LR-таблицы

6. Конечные преобразователи.

Понятие конечного преобразователя. Различные варианты его определения, их равносильность. Примеры преобразований, задаваемых конечными преобразователями. Теорема Нива об общем виде конечного преобразования. Замкнутость автоматных и контекстно-свободных языков относительно конечных преобразований. Теорема Хомского-Шютценберже, слабая и сильная формы.

7. Приложения теории формальных языков.

Современные приложения теории формальных языков. Применение конечных преобразователей в задачах автоматического морфологического анализа.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная мультимедиапроектором и экраном.

6. Перечень рекомендуемой литературы

Основная литература

1. Математическая теория формальных языков / А. Е. Пентус, М. Р. Пентус. — Москва, ИНТУИТ, 2016.— URL: <https://ibooks.ru/bookshelf/362911/reading> (дата обращения: 26.11.2020). - Полный текст (Режим доступа : из сети МФТИ / Удаленный доступ)

Дополнительная литература

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. J. Berstel. Transductions and context-free languages. — <http://www-igm.univ-mlv.fr/berstel/LivreTransductions/LivreTransductions14dec2009.pdf>
2. Вики-конспекты ИТМО. <https://neerc.ifmo.ru/wiki/index.php?title=%D0%A2%D0%B5%D0%BE%D1%80%D0%B8%D1%8F%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D1%85%D1%8F%D0%B7%D1%8B%D0%BA%D0%BE%D0%B2>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

- Генератор лексических анализаторов Flex (<http://flex.sourceforge.net/>),
- генератор синтаксических анализаторов Yacc/Bison (<http://www.gnu.org/software/bison/>),
- компилятор gcc/g++ для языка C++ (<https://gcc.gnu.org/>).

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий курс «Формальные языки и трансляции», должен как владеть теоретическим аппаратом, так и уметь решать практические задачи. Кроме того, студент должен уметь реализовывать основные алгоритмы, изучаемые в курсе, в виде программы на некотором языке программирования.

В результате освоения курса студент должен владеть основными понятиями данной предметной области, такими как формальный язык, конечный автомат, контекстно-свободная грамматика, уметь переходить от одних способов задания формальных языков к другим (например, от конечного автомата к регулярному выражению и обратно), а также применять данные понятия при решении теоретических и практических задач, в том числе программистского характера.

Изучение курса требует от студента напряжённой работы. Минимальное необходимое для этого количество часов приведено в таблице с содержанием курса. В самостоятельную работу входит:

- Посещение и конспектирование лекций и семинаров.
- Активное самостоятельное и коллективное решение задач на семинарах.
- Активное самостоятельное решение домашних задач.
- Проработку материалов лекций и семинаров, разбор утверждений и доказательств.
- Решение практических работ по курсу.
- Подготовка к экзамену.

Показателем владения материалом служит умение решать задачи. Для успешного освоения курса студенту следует разбирать все задачи, решаемые на семинарах или входящие в домашние задания. При решении задач необходимо осознавать, какие теоретические утверждения лежат в основе методов, применяемых в задачах.

При подготовке к семинарским занятиям следует повторять основные определения и формулировки теорем. Можно порекомендовать следующую схему освоения материала: повторение материала в день лекции (15-20 минут), повторение материала перед следующей лекцией (15-20 минут), проработка материала в течение недели по материалам семинарских и лекционных занятий (30-60 минут), решение домашних задач, выданных преподавателем (60 минут). Кроме того, для лучшего усвоения материала студенту рекомендуется индивидуально консультироваться с лектором/семинаристом по данному курсу.

Для промежуточного контроля материала используются 2 контрольные работы, а также 2 практических задания, предполагающие реализацию заданного алгоритма обработки регулярных выражений на некотором языке программирования, а также построение компилятора для фрагмента заданного языка программирования с помощью генератора синтаксических анализаторов.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладная математика и информатика
профиль подготовки:	АІ360: Передовые методы искусственного интеллекта Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
курс:	<u>3</u>
квалификация:	бакалавр
Семестр, формы промежуточной аттестации: 5 (осенний) - Экзамен	
Разработчик:	А.А. Сорокин, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности

2. Показатели оценивания компетенций

В результате изучения дисциплины «Формальные языки и трансляции» обучающийся должен:

знать:

- понятия формального языка и грамматики;
- классификацию формальных языков и грамматик;
- регулярные выражения;
- детерминированные и недетерминированные конечные автоматы;
- контекстно-свободные грамматики;
- основные алгоритмы синтаксического разбора для контекстно-свободных грамматик;
- автоматы с магазинной памятью;
- важнейшие подклассы контекстно-свободных грамматик (в т. ч. LR(k)-грамматики);
- конечные преобразователи
- практическое применение автоматов и грамматик, в т. ч. в задачах компиляции.

уметь:

делать правильные выводы из сопоставления результатов теории и эксперимента;
определять класс грамматик, необходимый для решения задачи;
формально доказывать соответствие формальных языков и грамматик поставленной задаче;
строить конечный автомат/регулярное выражение по описанию автоматного языка;
строить регулярное выражение по заданному конечному автомату;
строить конечный автомат по заданному регулярному выражению;
строить контекстно-свободную грамматику по описанию контекстно-свободного языка;
строить автомат с магазинной памятью для заданной контекстно-свободной грамматики;
строить синтаксический анализатор для заданной контекстно-свободной грамматики с помощью генераторов синтаксических анализаторов.

владеть:

математическим аппаратом теории формальных языков;
навыками самостоятельной работы в современных программных комплексах;
навыками освоения большого объема информации;
навыками использования генераторов синтаксических анализаторов.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

1. Построить регулярное выражение для языка слов, в которых число букв a чётно, а букв b – нечётно.
2. Построить конечный автомат для языка слов, в которых число букв a чётно, а букв b – нечётно.
3. Построить минимальный детерминированный конечный автомат для языка, задаваемого выражением $(a(ab+ba)^*b(a+ba)^*)^*$.
4. Найти регулярное выражение для дополнения языка $(a(ab+ba)^*b(a+ba)^*)$.
5. Проверить, является ли автоматным множество слов вида $anbncn$.
6. Построить контекстно-свободную грамматику для множества слов в алфавите $\{a,b,c\}$, где букв a больше, чем букв b .
7. Построить автомат с магазинной памятью для множества слов в алфавите $\{a,b,c\}$, где букв a больше, чем букв b .
8. Построить грамматику в нормальной форме Хомского для множества слов в алфавите $\{a,b,c\}$, где букв a больше, чем букв b .

9. Доказать, что язык слов вида $anbnc2n$ не является контекстно-свободным.
10. Проверить выводимость заданного слова в заданной грамматике с помощью алгоритма Кока-Янгера-Касами.
11. Проверить выводимость заданного слова в заданной грамматике с помощью алгоритма Эрли.
12. Построить LR-анализатор для заданной грамматики и проверить выводимость заданного слова с его помощью.
13. Построить конечный преобразователь, осуществляющий преобразование $a \rightarrow b \parallel a+c_d$.
14. Построить конечный преобразователь, умножающий число на 5 в двоичной системе счисления.

Контрольная работа состоит из 5 задач. Оценка «отлично» ставится, если обучающийся способен выбрать верный алгоритм во всех задачах, после чего довести их до корректного ответа. При этом решение должно быть полным и теоретически обоснованным. Оценка «хорошо» ставится, если не более, чем в одной задаче обучающийся выбрал неверный алгоритм, допустил ошибки в рассуждениях или обоснованиях. Оценка «удовлетворительно» ставится, если существенные недостатки имеются в двух или трёх задачах. Если число полностью решенных задач равно 1 и менее, ставится оценка «неудовлетворительно».

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

- Недетерминированные конечные автоматы (НКА). Различные варианты определений. Автоматные языки.
2. Детерминированные конечные автоматы (ДКА). Эквивалентность ДКА и НКА.
 3. Свойства класса автоматных языков. Замкнутость относительно булевых операций.
 4. Регулярные выражения. Теорема Клини о совпадении классов регулярных и автоматных языков.
 5. Лемма о разрастании для автоматных языков. Примеры неавтоматных языков.
 6. Минимальные ДКА. Алгоритм минимизации ДКА.
 7. Алгоритм проверки равенства регулярных выражений. Теорема Майхилла-Нерода.
 8. Праволинейные грамматики. Праволинейные языки. Теорема о совпадении классов автоматных и праволинейных языков.
 9. Контекстно-свободные грамматики (КС-грамматики). Контекстно-свободные языки (КС-языки). Вывод в КС-грамматике, деревья вывода. Однозначные КС-грамматики. Пример существенно-неоднозначного языка (без доказательства).
 10. Устранение бесполезных вспомогательных символов и ϵ -правил для КС-грамматик.
 11. Нормальная форма Хомского для КС-грамматик. Алгоритм приведения к нормальной форме Хомского.
 12. Алгоритм Кока-Янгера-Касами синтаксического разбора для КС-грамматик.
 13. Автоматы с магазинной памятью (МП-автоматы). Различные варианты определений. Языки, распознаваемые МП-автоматами.
 14. Совпадение классов КС-языков и языков, распознаваемых МП-автоматами.
 15. Лемма о разрастании для КС-языков. Примеры языков, не являющихся КС-языками.
 16. Алгоритм Эрли синтаксического разбора для КС-грамматик.
 17. Алгоритм синтаксического разбора по LR-таблице.
 18. Понятие LR-ситуации, состояния LR-анализатора. Операции GOTO и CLOSURE.
 19. Алгоритм построения LR-таблицы. Корректность LR-анализатора.
 20. Нормальная форма Грейбах для КС-грамматик.
 21. МП-автоматы без ϵ -переходов и распознаваемые ими языки.
 22. Детерминированные МП-автоматы. Детерминированные КС-языки. Различные варианты определения.
 23. Детерминированные КС-языки и однозначные КС-грамматики. Свойства класса детерминированных КС-языков (без доказательства).
 24. Конечные преобразователи и задаваемые ими преобразования. Различные

варианты определения. Примеры конечных преобразований. Теорема Нива.

25. Замкнутость конечных преобразований относительно композиции.

26. Замкнутость классов автоматных и КС-языков относительно конечных преобразований.

27. Лемма о разрастании для конечных преобразований. Примеры соответствий, не задаваемых конечными преобразованиями.

28. Теорема Хомского-Шютценберже.

29. Линейные и полуполулинейные множества. Теорема Парика.

Пример экзаменационного билета:

30. Минимальный детерминированный автомат. Его единственность.

31. Алгоритм Кока-Янгера-Касами. Доказательство корректности. Сложность алгоритма Кока-Янгера-Касами.

Пример экзаменационного билета:

1. Минимальный детерминированный автомат. Его единственность.

2. Алгоритм Кока-Янгера-Касами. Доказательство корректности. Сложность алгоритма Кока-Янгера-Касами.

Пример экзаменационной задачи:

Доказать, что в определении автомата с магазинной памятью можно не требовать очистки стека.

Пример задачи повышенной сложности:

Привести пример неавтоматного языка, удовлетворяющего лемме о разрастании для автоматного языка.

Критерии оценивания

Оценка «отлично» ставится, если обучающийся полно ответил на теоретический билет, привёл корректные доказательства всех требуемых утверждений, а также успешно ответил на дополнительные вопросы по программе, после чего решил 1-2 задачи повышенной сложности, возможно, с незначительной помощью экзаменатора.

Оценка «хорошо» ставится, если в теоретическом ответе имелись незначительные недостатки, после чего обучающийся справился лишь с задачами, требующими неложных рассуждений, но не смог решить задачу повышенной сложности.

Оценка «удовлетворительно» ставится при существенных пробелах в теоретических знаниях.

Если эти пробелы носят системный характер и касаются основных понятий курса, выставляется оценка «неудовлетворительно».

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Итоговая аттестация состоит из ответа по теоретическому билету, состоящему из 2 вопросов, а также решения нескольких задач, использующих теоретические знания. Количество задач зависит от качества ответа по теоретической части.