

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 25.08.2022 08:24:45
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»

Утверждаю:

Руководитель ООП:

С.М.Дудаков

2022 г.



Рабочая программа дисциплины (с аннотацией)
МЕТОДЫ МАТЕМАТИЧЕСКОЙ ЛИНГВИСТИКИ

Направление подготовки

09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

Профиль подготовки

Интеллектуальные системы. Теория и приложения

Для студентов 1 курса

Очная форма

Составитель: к.ф.-м.н. доцент Карлов Б.Н.

Тверь, 2022

I. Аннотация

1. Цель и задачи дисциплины:

Цель курса — ознакомить студентов с основными понятиями компьютерной лингвистики, с различными способами задания языков, с возможностью применения ЭВМ для обработки естественных языков.

2. Место дисциплины в структуре ООП

Дисциплина входит в раздел «Дисциплины профиля подготовки» части, формируемой участниками образовательных отношений, блока 1.

Предварительные знания и навыки. Знание курсов по математической логике, теории алгоритмов, теории автоматов и формальных языков.

Дальнейшее использование. Полученные знания используются для итоговой государственной аттестации, прохождении практики, а также в дальнейшей трудовой деятельности выпускников.

3. Объем дисциплины: 6 зач. ед., 216 акад. ч., в том числе:

контактная аудиторная работа лекций 16 ч., практических занятий 16 ч., **контактная внеаудиторная работа** контроль самостоятельной работы 0 ч., в том числе курсовая (расчетно-графическая) работа 0 ч.; **самостоятельная работа** 184 ч., в том числе контроль 36 ч.

4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы:

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
ПК-3, Способен выбирать и применять методы инженерии знаний для создания систем, основанных на знаниях	ПК-3.1, Выбирает и применяет методы сбора и извлечения знаний ПК-3.2, Выбирает и применяет методы структурирования знаний
ПК-10, Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых технологий искусственного интеллекта в прикладных областях	ПК-10.2, Руководит проектами в области сквозной цифровой технологии «Обработка естественного языка» ПК-10.4, Руководит проектами в области сквозной цифровой технологии «Распознавание и синтез речи»

5. Форма промежуточной аттестации и семестр прохождения:

экзамен во 2 семестре

6. Язык преподавания:

русский

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

Для студентов очной формы обучения

Учебная программа — наименование разделов и тем	Всего (час.)	Контактная работа (час.)				Сам. раб., в т.ч. контроль (час.)	
		Лекции		Практ. занятия / Лаб. работы			Контроль сам. раб., в т.ч. курсовая работа
		Всего	В т.ч. практ. подг.	Всего	В т.ч. практ. подг.		
1	2	3	4	5	6	7	8
Порождающие грамматики	36	2		2/0		0	32
Грамматики непосредственных составляющих	60	6		6/0		0	48
Категориальные грамматики	60	4		4/0		0	52
Слабо-контекстные грамматики	60	4		4/0		0	52
Итого	216	16	0	16/0	0/0	0	184

Учебная программа дисциплины

1. Порождающие грамматики

- Определение порождающих грамматик. Иерархия Хомского.
- Теорема о том, что грамматики типа 0 порождают в точности рекурсивно перечислимые языки.
- Теорема о представлении рекурсивно перечислимых языков.
- Алгоритмические проблемы для порождающих грамматик.

2. Грамматики непосредственных составляющих

- Теорема об эквивалентности НС-грамматик и неукорачивающих грамматик.
- Линейно ограниченные автоматы. Эквивалентность НС-грамматик и ЛО-автоматов.

- Нормальная форма Куроды.
- Свойства замкнутости для класса НС-языков.
- Пример рекурсивного языка, не являющегося НС-языком.
- Алгоритмические проблемы для НС-языков.
- PSPACE-полнота проблемы принадлежности для НС-языков.
- Деревья вывода и системы составляющих.

3. Категориальные грамматики

- Классические категориальные грамматики. Эквивалентность классических категориальных грамматик и КС-грамматик.
- Исчисление Ламбека и грамматики Ламбека.
- Эквивалентность грамматик Ламбека и КС-грамматик.
- Категориальные грамматики зависимостей (КГЗ). Основные свойства КГЗ.
- Алгоритм анализа для КГЗ.
- Деревья зависимостей. Связь деревьев зависимостей и систем составляющих.

4. Слабо-контекстные грамматики

- Индексные грамматики.
- Гнездовые стековые автоматы. Эквивалентность индексных грамматик и гнездовых автоматов.
- Множественные КС-грамматики.
- Линейные индексные грамматики.
- Комбинаторные категориальные грамматики.
- ТАГ-грамматики.
- Головные грамматики.
- Эквивалентность четырёх определений слабо-контекстных грамматик.
- Автоматы со вложенными магазинами. Их эквивалентность ТАГ-грамматикам.
- Применения комбинаторных категориальных грамматик для представления семантики в виде λ -термов.

III. Образовательные технологии

Учебная программа — наименование разделов и тем	Вид занятия	Образовательные технологии
---	-------------	----------------------------

Учебная программа — наименование разделов и тем	Вид занятия	Образовательные технологии
Порождающие грамматики	лекции, практические занятия	изложение теоретического материала, решение задач
Грамматики непосредственных составляющих	лекции, практические занятия	изложение теоретического материала, решение задач
Категориальные грамматики	лекции, практические занятия	изложение теоретического материала, решение задач
Слабо-контекстные грамматики	лекции, практические занятия	изложение теоретического материала, решение задач

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Типовые контрольные задания и/или критерии для проверки индикатора ПК-3.1

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить различные слабо-контекстные грамматики и автоматы по описанию языка	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> Постройте ТАГ-грамматику для языка $L = \{ a^i b^j c^k \mid 0 < i < j < k \}.$ Постройте комбинаторную категориальную грамматику для языка $L = \{ w * w^{-1} * w \mid w \in \{0, 1\}^* \}.$ Постройте автомат с вложенными магазинами для языка $L = \{ a^{3i} b^{2i} c^i \mid i > 0 \}.$ Постройте гнездовой автомат для языка $L = \{ a^{2^n} \mid n > 0 \}.$ 	оценка 3 — умеет строить слабо-контекстные грамматики по описанию языка, оценка 4 — кроме того умеет строить автоматы с вложенными магазинами по описанию языка, оценка 5 — кроме того умеет строить гнездовые автоматы по описанию языка

Типовые контрольные задания и/или критерии для проверки индикатора ПК-3.2

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Уметь строить неукорачивающие грамматики и линейно-ограниченные автоматы по описанию языка	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> Постройте неукорачивающую грамматику, порождающую следующий язык: $L = \{ a^i b^j c^k \mid 0 < i \leq j, j > k > 0 \}.$ Постройте линейно-ограниченный автомат, распознающий 	оценка 3 — умеет строить неукорачивающие грамматики по описанию языка, оценка 4 — кроме того умеет строить линейно-ограниченные автоматы по описанию

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<p>следующий язык:</p> $L = \{ a^{3^n} \mid n > 0 \}.$ <p>Примеры тем для самостоятельной работы:</p> <ul style="list-style-type: none"> • Напишите программу, которая по НС-грамматике и предложению на русском языке строит все его размеченные системы составляющих. 	<p>языка, оценка 5 — кроме того умеет обосновывать правильность построения</p>
<p>Знать понятие категориальной грамматики, дерева зависимостей, связь деревьев зависимостей с категориальными грамматиками</p>	<ul style="list-style-type: none"> • Классические категориальные грамматики. Эквивалентность классических категориальных грамматик и КС-грамматик. • Исчисление Ламбека и грамматики Ламбека. • Эквивалентность грамматик Ламбека и КС-грамматик. • Категориальные грамматики зависимостей (КГЗ). Основные свойства КГЗ. • Алгоритм анализа для КГЗ. • Деревья зависимостей. Проективные и непроективные деревья. Построение деревьев зависимостей с помощью категориальных грамматик. 	<p>оценка 3 — знает понятия классической категориальной грамматики, дерева зависимостей, оценка 4 — кроме того знает другие типы категориальных грамматик, оценка 5 — кроме того знает доказательства перечисленных утверждений</p>
<p>Уметь строить категориальные грамматики по описанию языка, строить деревья зависимостей</p>	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • Постройте классическую категориальную грамматику для языка $L = \{ a^i b^j c^k \mid i, j, k \geq 0, i = j \text{ или } j = k \}.$ • Постройте категориальную грамматику зависимостей для языка $L = \{ w \in \{ a, b, c, d \}^+ \mid w _a + w _c \geq w _b + w _d \text{ и все символы } a \text{ стоят правее } c \}.$ • Докажите в исчислении Ламбека L секвенцию $p/(q \setminus (r \cdot (r \setminus q))) \rightarrow (p/(r \setminus q))/(q \setminus r).$ <p>Примеры тем для самостоятельной работы:</p> <ul style="list-style-type: none"> • Напишите программу, которая по классической категориальной грамматике и предложению на русском языке строит все его размеченные деревья зависимостей. 	<p>оценка 3 — кроме того умеет строить категориальные грамматики по описанию языка, оценка 4 — кроме того умеет обосновывать правильность построения, оценка 5 — кроме того умеет выводить секвенции в исчислении Ламбека</p>
<p>Уметь строить формальное представление синтаксиса и семантики предложений на естественных языках</p>	<p>Примеры задач для контрольных работ:</p> <ul style="list-style-type: none"> • Постройте размеченную иерархизованную систему составляющих для предложения «К востоку от боровых озёр лежат громадные мещёрские болота — мшары». • Постройте размеченное дерево зависимостей для предложения «Для точной диагностики заболеваний внутренних органов человека рентген незаменим». • Дано предложение «Мастер внимательно осматривал станок». Его словам сопоставлены следующие категории и λ-термы: <ul style="list-style-type: none"> – мастер — NP : M – внимательно — C : V 	<p>оценка 3 — умеет строить системы составляющих и деревья зависимостей для предложений, оценка 4 — кроме того умеет строить размеченные иерархизованные системы составляющих и размеченные деревья зависимостей, оценка 5 — кроме того умеет представ-</p>

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<p>– осматривал – $((S \setminus NP) \setminus C) / NP : \lambda xyz.Ozxy$</p> <p>– станок – $NP : C$</p> <p>Сократите категории до S и упростите получающийся λ-терм. На первом шаге примените к первой категории правило ($> T$), а на втором шаге примените к третьей и четвертой категориям правило ($>$).</p>	<p>лять семантику в виде λ-термов</p>

Типовые контрольные задания и/или критерии для проверки индикатора ПК-10.2

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
<p>Знать понятия порождающей грамматики, системы составляющих, связь систем составляющих с грамматиками</p>	<ul style="list-style-type: none"> • Определение порождающих грамматик. Иерархия Хомского. • Теорема о том, что грамматики типа 0 порождают в точности рекурсивно перечислимые языки. • Теорема о представлении рекурсивно перечислимых языков. • Алгоритмические проблемы для порождающих грамматик. • Грамматики непосредственных составляющих. Эквивалентность НС-грамматик и неукорачивающих грамматик. • Линейно ограниченные автоматы. Эквивалентность НС-грамматик и ЛО-автоматов. • Нормальная форма Куроды. • Свойства замкнутости для класса НС-языков. • Пример рекурсивного языка, не являющегося НС-языком. • Алгоритмические проблемы для НС-языков. • PSPACE-полнота проблемы принадлежности для НС-языков. • Системы составляющих, размеченные системы составляющих, иерархизованные системы составляющих. Построение системы составляющих по дереву вывода. 	<p>оценка 3 — знает определения основных понятий (НС-грамматика, ЛО-автомат, система составляющих), оценка 4 — кроме того знает основные свойства НС-грамматик, оценка 5 — кроме того знает доказательства перечисленных утверждений</p>

Типовые контрольные задания и/или критерии для проверки индикатора ПК-10.4

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
<p>Знать понятие слабо-контекстных грамматик</p>	<ul style="list-style-type: none"> • Индексные грамматики. • Гнездовые стековые автоматы. Эквивалентность индексных грамматик и гнездовых автоматов. • Понятие слабо-контекстных грамматик. • Множественные КС-грамматики. 	<p>оценка 3 — знает различные классы слабо-контекстных грамматик, оценка 4 — кроме того знает различные свойства слабо-контекстных грамматик, оценка</p>

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	<ul style="list-style-type: none"> • Линейные индексные грамматики. Комбинаторные категориальные грамматики. ТАГ-грамматики. Головные грамматики. • Эквивалентность четырёх определений слабо-контекстных грамматик. • Автоматы со вложенными магазинами. Их эквивалентность ТАГ-грамматикам. • λ-термы. Применения комбинаторных категориальных грамматик для представления семантики в виде λ-термов. 	5 — кроме того знает доказательства перечисленных утверждений

V. Учебно-методическое и информационное обеспечение дисциплины

1. Рекомендованная литература

а) Основная литература

- [1] Волосатова, Т.М. Информатика и лингвистика [Электронный ресурс] : Учебное пособие / Волосатова Т.М., Чичварин Н.В. — Электрон. дан. — М.: НИЦ ИНФРА-М, 2016. — 196 с.: 60x90 1/16. — (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0 — Режим доступа: <http://znanium.com/bookread2.php?book=508097>
- [2] Марченков, С.С. Конечные автоматы [Электронный ресурс] : учеб. пособие — Электрон. дан. — Москва : Физматлит, 2008. — 56 с. — Режим доступа: <https://e.lanbook.com/book/59510>. — Загл. с экрана.
- [3] Короткова, М.А. Задачник по курсу "Математическая лингвистика и теория автоматов": учебное пособие для вузов [Электронный ресурс] : учеб. пособие / М.А. Короткова, Е.Е. Трифонова. — Электрон. дан. — Москва : НИЯУ МИФИ, 2012. — 92 с. — Режим доступа: <https://e.lanbook.com/book/75843>. — Загл. с экрана.

б) Дополнительная литература

- [4] Федосеева, Л.И. Основы теории конечных автоматов и формальных языков [Электронный ресурс] : учеб. пособие / Л.И. Федосеева, Р.М. Адилов, М.Н. Шмокин. — Электрон. дан. — Пенза : ПензГТУ, 2013. — 136 с. — Режим доступа: <https://e.lanbook.com/book/62703>. — Загл. с экрана.

2. Программное обеспечение

Наименование помещений	Программное обеспечение
Ауд. 308 (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Google Chrome — бесплатно; Microsoft Office 365 pro plus — Акт на передачу прав № 687 от 31 июля 2018; Microsoft Windows 10 Enterprise — Акт на передачу прав № 687 от 31 июля 2018; Kaspersky Endpoint Security 10 для Windows — Акт на передачу прав №2129 от 25 октября 2016 г.
Ауд. 201а (компьютерная лаборатория ПМиК) (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Перечень программного обеспечения (со свободными лицензиями): Linux OpenSuse Tumbleweed, KDE, TeXLive, Mozilla Firefox, TeXStudio, Qt, QtCreator, Gcc, Python, Eric, LibreOffice, Cervisia, Kdbg, Umbrello, wxMaxima, Blender, digikam, GIMP, Gwenview, hugin, Inkscape, Okular, showFoto, Kmail, Konqueror, Konversation, Kopete, TigerVNC viewer, Amarok, K3b, Kdenlive, VLC media player, Kontact, Korganizer, Yast, Ark, Dolphin, Info Center, Kget, Konsole, Krusader, Midnight commander, OpenJDK, pgadmin3, Xterm, Emacs, Kate, Kcalc, Kpgp, Kleopatra, Kompare, Sweeper, Perl, Apache, PostgreSQL, MariaDB, SQLite, PHP

3. Современные профессиональные базы данных и информационные справочные системы

- [1] ЭБС «ZNANIUM.COM» <http://www.znanium.com>
- [2] ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru>
- [3] ЭБС IPRbooks <http://www.iprbookshop.ru>
- [4] ЭБС «Лань» <http://e.lanbook.com>
- [5] ЭБС «Юрайт» <https://urait.ru>
- [6] ЭБС ТвГУ <http://megapro.tversu.ru/megapro/Web>
- [7] Научная электронная библиотека eLIBRARY.RU (подписка на журналы) https://elibrary.ru/projects/subscription/rus_titles_open.asp
- [8] Репозиторий ТвГУ <http://eprints.tversu.ru>

4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

- [1] Natural Language Processing, <http://www.learnerstv.com/Free-Computer-Science-V>
- [2] Московский центр непрерывного математического образования, <http://www.mccme.ru/>

VI. Методические материалы для обучающихся по освоению дисциплины

Примеры задач для подготовки к контрольным работам

1. Постройте гнездовой стековый автомат для языка $L = \{a^{n^2} \mid n > 0\}$ и обоснуйте его правильность.
2. Постройте комбинаторную категориальную грамматику для языка $L = \{w\#w^{-1} \mid w - \text{правильное скобочное слово}\}$ и обоснуйте ее правильность.
3. Постройте автомат с вложенными магазинами для языка $L = \{a^n b^n c^n d^n \mid n > 0\}$.
4. Постройте НС-грамматику для языка $L = \{a^n \mid n - \text{составное число}\}$.
5. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Тощая торговка вяленой воблой торчала среди ящиков».
6. Обобщенной НС-грамматикой (ОНС-грамматикой) называется порождающая грамматики $G = \langle \Sigma, N, S, P \rangle$, правила которой имеют вид $\xi A \eta \rightarrow \xi \alpha \eta$, где $\xi, \eta, \alpha \in (\Sigma \cup N)^*$, $A \in N$. Докажите, что любой рекурсивно перечислимый язык порождается некоторой ОНС-грамматикой.
7. Докажите, что класс языков типа 0 замкнут относительно тасовки:

$$\text{TAC}(L_1, L_2) = \bigcup_{x \in L_1, y \in L_2} \{x_1 y_1 \dots x_n y_n \mid x_i, y_j \in \Sigma^*, x_1 \dots x_n = x, y_1 \dots y_n = y\}.$$

Требования к рейтинг контролю (7 семестр)

Контрольная работа 1. Темы: порождающие грамматики, НС-грамматики, системы составляющих, деревья зависимостей. Пример задания:

1. Постройте неукорачивающую грамматику, порождающую следующий язык:

$$L = \{a^n b^m a^n b^m \mid m, n > 0\}.$$

2. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Лебеди стадом летели из холодной стороны в тёплые земли».
3. Докажите, что класс языков типа 0 замкнут относительно правого деления:

$$L_1/L_2 = \{w \in \Sigma^* \mid \text{существует слово } x \in L_2 \text{ такое, что } wx \in L_1\}.$$

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 1. Темы: НС-грамматики, системы составляющих, деревья зависимостей. Пример задания:

Напишите программу, которая по системе составляющих строит согласованное с ней дерево зависимостей.

За решение задачи выставляется максимум 20 баллов.

Контрольная работа 2. Темы: категориальные грамматики, слабо-контекстные грамматики. Пример задания:

1. Постройте линейную индексную грамматику для языка

$$L = \{ w * w^{-1} * w \mid w \in \{0, 1\}^* \}.$$

2. Постройте гнездовой автомат для языка

$$L = \{ a^n b^{n^2+m} \mid n \geq 0, m \geq 0 \}.$$

3. Разрешим использовать в гнездовых автоматах ещё три типа команд: L_x — перейти к левому брату, если метка ребра, входящего в текущую вершину, равна x ; P_x — перейти к правому брату, если метка ребра, входящего в текущую вершину, равна x ; $I_{x,y}$ — изменить метку текущего ребра с x на y . Докажите, что любой язык типа 0 распознаётся расширенным гнездовым автоматом.

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 2. Темы: слабо-контекстные грамматики. Пример задания:

Напишите программу, которая по линейной индексной грамматике и слову проверяет, выводимо ли слово в этой грамматике.

За решение задачи выставляется максимум 20 баллов.

Общая сумма В сумме за все задачи выставляет не более 100 баллов.

VII. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Для аудиторной работы

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 308 (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Ауд. 308 приспособлена для проведения лекционных и практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации и оснащена набором учебной мебели, меловой доской, настенным экраном (экран на треноге Da-lite versatal 213x213)) и проектором Samsung SP D300BX.

Для самостоятельной работы

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 201а (компьютерная лаборатория ПМиК) (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Ауд. 201а (компьютерная лаборатория ПМиК) оснащена персональными ЭВМ (компьютер ПЭВМ «ХОПЕР» IS09001: 1.1/Intel Core i3-540/IntelH55-MLX/Unix-11.4/DVD RW Sony/Монитор 21,5” AOC TFT/клавиатура/мышь — 10 штук) с доступом к сети Интернет и необходимым программным обеспечением, системным блоком BASE P4 3200MHz 800 512K/1024 Мб DDR400/400Gb, концентратором сетевым DFE-916 DX HUB 16x10/100.

VIII. Сведения об обновлении рабочей программы дисциплины

№ п/п	Обновленный раздел рабочей программы дисциплины	Описание внесённых изменений	Дата и протокол заседания кафедры, утвердившего изменения
1	I. Аннотация. 3. Объем дисциплины	Выделение часов на практическую подготовку	От 29.10.2020 года, протокол №3 учёного совета факультета
2	II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	Выделение часов на практическую подготовку по темам	От 29.10.2020 года, протокол №3 учёного совета факультета
3	I. Аннотация. IV. Оценочные материалы для проведения текущей и промежуточной аттестации	Изменения в учебные планы и в рабочие программы дисциплин, формирующих новые/измененные компетенции в соответствии с приказом Минобрнауки России от 26.11.2020 г. №1456	Решение научно-методического совета (протокол №6 от 02.06.2021 г.)